



## NEISSERIAL POLYNUCLEOTIDES

This application is a continuation-in-part of international patent application PCT/IB98/01665, filed October 9, 1998, from which priority is claimed under 35 U.S.C. § 119.

This invention relates to antigens from *Neisseria* bacteria.

### 5 BACKGROUND ART

*Neisseria meningitidis* and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely  
10 different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

*N.gonorrhoeae* caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In:  
15 *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1  
20 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

*N.meningitidis* causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis*  
25 Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during

epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

- 5 Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a
- 10 tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [*eg. Morbidity and Mortality weekly report*, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the
- 15 vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).
- 20 Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of  $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and
- 25 therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular. *Clin Microbiol Rev* 7(4):559-575).



Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (*eg.* Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (*eg.* Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (*eg.* EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

## THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (*eg.* 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the

Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

5 The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

10 The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

15 According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide sequences disclosed in the examples.

20 Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, *n* is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

25 According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various  
5 forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

10 According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*eg.* as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid,  
15 protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B  
20 or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

25 A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing  
5 conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the  
10 invention (*eg.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

### General

The practice of the present invention will employ, unless otherwise indicated, conventional  
15 techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J.  
20 Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology*  
25 (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

## 5 Definitions

A composition containing X is “substantially free of” Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term “comprising” means “including” as well as “consisting” *eg.* a composition “comprising”

10 X may consist exclusively of X or may include something additional to X, such as X+Y.

A “conserved” *Neisseria* amino acid fragment or protein is one that is present in a particular Neisserial protein in at least x% of *Neisseria*. The value of x may be 50% or more, e.g., 66%, 75%, 80%, 90%, 95% or even 100% (i.e. the amino acid is found in the protein in question in all *Neisseria*). In order to determine whether an amino acid is “conserved” in a particular Neisserial

15 protein, it is necessary to compare that amino acid residue in the sequences of the protein in question from a plurality of different *Neisseria* (a reference population). The reference population may include a number of different *Neisseria* species or may include a single species. The reference population may include a number of different serogroups of a particular species or a single serogroup. A preferred reference population consists of the 5 most common *Neisseria*. The term

20 “heterologous” refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two

25 epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain  
5 origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having  
10 sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid  
15 molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of  
20 the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

### Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

#### i. Mammalian Systems

25 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding

sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element  
5 determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences.  
10 Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

15 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped  
20 orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus  
25 [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as



mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.

## ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the

heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human  $\alpha$ -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired,

methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein  
5 comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector  
10 and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion  
15 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

20 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein,  
25 which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15  $\mu$ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells

infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg.* Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence.

These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

### iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art.

- 5 Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 10 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant* 15 *Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

- Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. 25 Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A

general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed

out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*,



*Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.*

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue  
5 is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop  
10 simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the  
15 invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be  
20 adjusted through routine methods to optimize expression and recovery of heterologous protein.

#### iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation  
25 region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits

negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid

promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9  
5 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and*  
10 *Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a  
15 methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end  
20 of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from  
25 the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that

preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an

extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number  
5 ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an  
10 integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of  
15 bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes,  
20 such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

25 Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia*

coli [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], Streptomyces lividans [US patent  
5 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl<sub>2</sub> or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.*  
10 [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, Campylobacter], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived  
15 plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus]; [Barany *et al.* (1980) *J. Bacteriol.*  
20 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, Streptococcus].

#### v. Yeast Expression

25 Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the

“TATA Box”) and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) “The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*,” in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with  
5 cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two  
10 amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific  
15 processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites  
20 encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US  
25 patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor



fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (eg. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCl/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences

flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*,  
5 *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable  
10 integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to  
15 tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation  
20 vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol.*  
25 *Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.*

(1990) *Bio/Technology* 8:135], *Pichia guillerimondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse  
5 (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];  
10 [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos.  
15 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

### Antibodies

20 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised  
25 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200  $\mu$ g/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (*eg.* 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*eg.* hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (*eg.* in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly  $^{32}\text{P}$  and  $^{125}\text{I}$ ), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example,  $^{125}\text{I}$  may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with  $^{125}\text{I}$ , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

### Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount

in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hypodermic sprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

10 Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with “pharmaceutically acceptable carriers,” which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents (“adjuvants”). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, *etc.*; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds.

Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi<sup>TM</sup> adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox<sup>TM</sup>);  
10 (3) saponin adjuvants, such as Stimulon<sup>TM</sup> (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (*eg.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*eg.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), *etc.*; and (6) other  
15 substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59<sup>TM</sup> are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-  
20 hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.  
25 Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.



Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

#### Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral,

adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *eg.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly

preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5

native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC

VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example

ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex

beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

- 5 Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like
- 10 polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci.*
- 15 *USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033
- 20 Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.
- 25 A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects  
5 can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or  
10 transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic  
15 cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well  
20 known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

25 One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating



factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

5    B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, etc.

10    Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

15    The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim.*  
20    *Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified  
25    transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE  
5 (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids  
10 (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the  
15 art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta*  
20 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

## 25 E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally

occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

- 5 Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

- 10 A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

- The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) Annu Rev. Biochem 54:699; Law (1986) Adv. Exp Med. Biol. 151:162; Chen (1986) J Biol Chem  
15 261:12918; Kane (1980) Proc Natl Acad Sci USA 77:2465; and Utermann (1984) Hum Genet 65:232.

- Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of  
20 naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

- Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance.  
25 Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for

example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

## 5 F.Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired  
10 location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from  
15 DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

20 The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when  
25 combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods.

- 5 Antibodies to Neisserial proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody  
10 or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

- Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed  
15 by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

- “Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen  
20 bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt’s reagent or BLOTTO); concentration  
25 of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated  $T_m$  of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to  $10^{-9}$  to  $10^{-8}$  g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of  $10^8$  cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than  $10^8$  cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature ( $T_m$ ) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where  $C_i$  is the salt concentration (monovalent ions) and  $n$  is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

## 20 Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

25 The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be

complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].



Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence  
5 that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are  
10 generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified  
15 and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

## BRIEF DESCRIPTION OF THE DRAWINGS

20 **Figures 1-20** show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37 (Fig. 1A-1E), 5 (Fig. 2A-2B), 2 (Fig. 3A-3D), 15 (Fig. 4A-4C), 22 (Fig. 5A-5C), 28 (Fig. 6A-6B), 32 (Fig. 7A-7B), 4 (Fig. 8A-8F), 61 (Fig. 9), 76 (Fig. 10A-10C), 89 (Fig. 11), 97 (Fig. 12A-12E), 106 (Fig. 13A-7C), 138 (Fig. 14A-B), 23 (Fig. 15A-15C), 25 (Fig. 16A-16E), 27 (Fig. 17A-17B), 79 (Fig. 18A-18B), 85 (Fig. 19A-19D) and 132 (Fig. 20A-20C). M1 and M2 are  
25 molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (◆) shows preimmune data; a triangle (▲)

shows GST control data; a circle ( ) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

**Figure 21** shows an alignment comparison of amino acid sequences for ORF 4 for several strains of *Neisseria*. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. The Figure demonstrates a high degree of conservation among the various strains, further confirming its utility as an antigen for both vaccines and diagnostics.

## EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of

known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also  
5 Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as  
10 implemented at [http://www.genome.ou.edu/gono\\_blast.html](http://www.genome.ou.edu/gono_blast.html). The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID NO: 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID NO: 11)  
15 represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical  
20 evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in  
25 the ORFs, as predicted by the PSORT algorithm (<http://www.psort.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

#### A) Chromosomal DNA preparation

*N.meningitidis* strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one ChCl<sub>3</sub>/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

#### B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems:

5 pGEX-KG (using either *Bam*HI-*Xho*I or *Eco*RI-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

5'-end primer tail: CGCGGATCCCATATG (SEQ ID NO: 1099) (*Bam*HI-*Nde*I)

CGCGGATCCGCTAGC (SEQ ID NO: 1100) (*Bam*HI-*Nhe*I)

CCGGAATTCTAGCTAGC (SEQ ID NO: 1101) (*Eco*RI-*Nhe*I)

10 3'-end primer tail: CCCGCTCGAG (SEQ ID NO: 1102) (*Xho*I)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (SEQ ID NO: 1103) (*Nde*I)

15 5'-end primer tail: CGGGATCC (*Bam*HI)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *Nhe*I-*Bam*HI restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (SEQ ID NO: 1104) (*Nhe*I)

20 3'-end primer tail: CGGGATCC (*Bam*HI)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

25  $T_m = 4 (G+C) + 2 (A+T)$  (tail excluded)

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N$$

(whole primer)

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

**Table I** shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed:

ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

**TABLE I – PCR primers**

ORF	Primer	Sequence	Restriction sites
<b>ORF 1</b>	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG (SEQ ID NO: 924)	BamHI-NheI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT (SEQ ID NO: 925)	XhoI
<b>ORF 2</b>	Forward	GCGGATCCCATATG-TTTGATTTTCGGTTTGGG (SEQ ID NO: 926)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG (SEQ ID NO: 927)	XhoI
<b>ORF 2-1</b>	Forward	GCGGATCCCATATG-TTTGATTTTCGGTTTGGG (SEQ ID NO: 928)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGATTTACGGACGCGCA (SEQ ID NO: 929)	XhoI
<b>ORF 4</b>	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC (SEQ ID NO: 930)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC (SEQ ID NO: 931)	XhoI
<b>ORF 5</b>	Forward	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAACC (SEQ ID NO: 932)	NdeI-NcoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAC (SEQ ID NO: 933)	BamHI

ORF 6	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG (SEQ ID NO: 934)	XhoI
	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA (SEQ ID NO: 935)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC (SEQ ID NO: 936)	XhoI
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTCC (SEQ ID NO: 937)	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTTGCGGA (SEQ ID NO: 938)	XhoI
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC (SEQ ID NO: 939)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC (SEQ ID NO: 940)	XhoI
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA (SEQ ID NO: 941)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCGAGGTTTTCGGG (SEQ ID NO: 942)	XhoI
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC (SEQ ID NO: 943)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAATGGGAAACCTTGTTTT (SEQ ID NO: 944)	XhoI
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG (SEQ ID NO: 945)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC (SEQ ID NO: 946)	XhoI
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC (SEQ ID NO: 947)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG (SEQ ID NO: 948)	XhoI
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG (SEQ ID NO: 949)	NdeI-NcoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG (SEQ ID NO: 950)	BamHI
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTATG (SEQ ID NO: 951)	XhoI
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTCG (SEQ ID NO: 952)	NdeI-NcoI
	Forward	CGGGATCC-ATTGCCGGCCTGTTCG (SEQ ID NO: 953)	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC (SEQ ID NO: 954)	XhoI
ORF 18	Forward	GCGGATCCCATATG-ATTTTGCTGCATTTGGAT (SEQ ID NO: 955)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC (SEQ ID NO: 956)	XhoI

ORF 19	Forward	GGAATTCCATATGGCCATGG -TCGCCAGTGTTTTTACC (SEQ ID NO: 957)	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGTTTTTACCG (SEQ ID NO: 958)	BamHI
	Reverse	CCCGCTCGAG-GGTGTTTTTGAAGCTGCC (SEQ ID NO: 959)	XhoI
ORF 20	Forward	GGAATTCCATATGGCCATGG -TCGGCGCGGGTATG (SEQ ID NO: 960)	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGCGGGTATG (SEQ ID NO: 961)	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA (SEQ ID NO: 962)	XhoI
ORF 22	Forward	GGAATTCCATATGGCCATGG-TGATTAAATCAAAAAGGTCT (SEQ ID NO: 963)	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAATCAAAAAGGTCTAAACC (SEQ ID NO: 964)	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCC (SEQ ID NO: 965)	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC (SEQ ID NO: 966)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG (SEQ ID NO: 967)	XhoI
ORF 24	Forward	GGAATTCCATATGGCCATGG -TGATGCCGGAATGGTG (SEQ ID NO: 968)	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAATGGTG (SEQ ID NO: 969)	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGGCGCA (SEQ ID NO: 970)	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGCAAATGATTGC (SEQ ID NO: 971)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG (SEQ ID NO: 972)	XhoI
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC (SEQ ID NO: 973)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGGTTTT (SEQ ID NO: 974)	XhoI
ORF 27	Forward	GGAATTCCATATGGCCATGG-AGACCTATTCTGTTTA (SEQ ID NO: 1168)	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC (SEQ ID NO: 975)	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT (SEQ ID NO: 976)	XhoI
ORF 28	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT (SEQ ID NO: 977)	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG (SEQ ID NO: 978)	BamHI



<b>ORF 29</b>	Reverse	978) CCCGCTCGAG-TTTGTCAGAGGAATTCGCG (SEQ ID NO: 979)	XhoI
	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCCG (SEQ ID NO: 980)	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC -AACGGTTTGGATGCCCCG (SEQ ID NO: 981)	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTGTCTAAGTTCCTGATATG (SEQ ID NO: 982)	XhoI
<b>ORF 32</b>	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG (SEQ ID NO: 983)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTTTGATGCTTTG (SEQ ID NO: 984)	XhoI
<b>ORF 33</b>	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG (SEQ ID NO: 985)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC (SEQ ID NO: 986)	XhoI
<b>ORF 35</b>	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT (SEQ ID NO: 987)	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT (SEQ ID NO: 988)	BamHI-NheI
	Reverse	CCCGCTCGAG-AAACAGCCATTTGAGCGA (SEQ ID NO: 989)	XhoI
<b>ORF 37</b>	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT (SEQ ID NO: 990)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAGCCCGCTTTCAGG (SEQ ID NO: 991)	XhoI
<b>ORF 58</b>	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT (SEQ ID NO: 992)	BamHI-NheI
	Reverse	CCCGCTCGAG-AGCATTGTCCAAGGGGAC (SEQ ID NO: 993)	XhoI
<b>ORF 65</b>	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG (SEQ ID NO: 994)	NdeI-NcoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG (SEQ ID NO: 995)	BamHI
	Reverse	CCCGCTCGAG-CCGCATCGGCAGACA (SEQ ID NO: 996)	XhoI
<b>ORF 66</b>	Forward	GCGGATCCCATATG-TACGCATTTACCGCCG (SEQ ID NO: 997)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TGGATTTGCAGAGATGG (SEQ ID NO: 998)	XhoI
<b>ORF 72</b>	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA (SEQ ID NO: 999)	BamHI-NdeI

ORF 73	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA (SEQ ID NO: 1000)	XhoI
	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG (SEQ ID NO: 1001)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCATCTTTTTCATGTTG (SEQ ID NO: 1002)	XhoI
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC (SEQ ID NO: 1003)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG (SEQ ID NO: 1004)	XhoI
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC (SEQ ID NO: 1005)	NheI-NdeI
	Reverse	CGGGATCC-TTACGGTTTGACACCGTT (SEQ ID NO: 1006)	BamHI
ORF 79	Forward	CGCGGATCCCATATG-GTTTCCGCCGCCG (SEQ ID NO: 1007)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG (SEQ ID NO: 1008)	XhoI
ORF 83	Forward	GCGGATCCCATATG-AAAACCCTGCTGCTGC (SEQ ID NO: 1009)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCCGCCTTTGCGGC (SEQ ID NO: 1010)	XhoI
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG (SEQ ID NO: 1011)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA (SEQ ID NO: 1012)	XhoI
ORF 85	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA (SEQ ID NO: 1013)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCGGCGCGGCGGGC (SEQ ID NO: 1014)	XhoI
ORF 89	Forward	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA (SEQ ID NO: 1015)	NdeI-NcoI
	Forward	CGGGATCC-GCCATACCTTCTTATCAGAG (SEQ ID NO: 1016)	BamHI
	Reverse	CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC (SEQ ID NO: 1017)	XhoI
ORF 97	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC (SEQ ID NO: 1018)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG (SEQ ID NO: 1019)	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGGTAACTGCGG (SEQ ID NO: 1020)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTGTTGGGGCAAATC (SEQ ID NO: 1021)	XhoI

ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTTCACACCG (SEQ ID NO: 1022)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGAA (SEQ ID NO: 1023)	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAAGAAACCTC (SEQ ID NO: 1024)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTTCCGCCTTTCAATGT (SEQ ID NO: 1025)	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC (SEQ ID NO: 1026)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC (SEQ ID NO: 1027)	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC (SEQ ID NO: 1028)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC (SEQ ID NO: 1029)	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC (SEQ ID NO: 1030)	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC (SEQ ID NO: 1031)	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC (SEQ ID NO: 1032)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG (SEQ ID NO: 1033)	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG (SEQ ID NO: 1034)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTGTTCCCGATGATGTT (SEQ ID NO: 1035)	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG (SEQ ID NO: 1036)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG (SEQ ID NO: 1037)	XhoI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC (SEQ ID NO: 1038)	EcoRI
	Reverse	AAACTGCAG-GGAAAACCATCCGCACTCTGCC (SEQ ID NO: 1039)	PstI
ORF111	Forward	AAAGAATTC-GCACCGCAAAGGCAAAAACCGCA (SEQ ID NO: 1040)	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGTTTTCGGGCAGGGTGG (SEQ ID NO: 1041)	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG (SEQ ID NO: 1042)	EcoRI

ORF115	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG (SEQ ID NO: 1043)	PstI
	Forward	AAAGAATTC-TTGCTTGTGCAAACAGAAAAGACGG (SEQ ID NO: 1044)	EcoRI
	Reverse	AAAAAAGTCGAC- CTATTTTTTAGGGGCTTTGCTTGTGAAAAGCCTGCC (SEQ ID NO: 1045)	Sall
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG (SEQ ID NO: 1046)	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC (SEQ ID NO: 1047)	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG (SEQ ID NO: 1048)	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT (SEQ ID NO: 1049)	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC (SEQ ID NO: 1050)	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC (SEQ ID NO: 1051)	PstI
ORF122	Forward	AAAAAAGTCGAC-ATGTCTTACCGCGCAAGCAGTTCTCC (SEQ ID NO: 1052)	Sall
	Reverse	AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC (SEQ ID NO: 1053)	PstI
ORF125	Forward	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT (SEQ ID NO: 1054)	EcoRI
	Reverse	AAACTGCAG-CGCCGTTTCAAGACGAAAAGTCG (SEQ ID NO: 1055)	PstI
ORF126	Forward	AAAGAATTC-GCGGAAACGGTCAAG (SEQ ID NO: 1056)	EcoRI
	Reverse	AAACTGCAG-TTAATCTTGTCTTCCGATATAC (SEQ ID NO: 1057)	PstI
ORF127	Forward	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG (SEQ ID NO: 1058)	EcoRI
	Reverse	AAAAAAGTCGAC-CTTAAGTAACTGCAGTCCTTATC (SEQ ID NO: 1059)	Sall
ORF128	Forward	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC (SEQ ID NO: 1060)	EcoRI
	Reverse	AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTGAGCAGGC G (SEQ ID NO: 1061)	PstI
ORF129	Forward	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATAC CG (SEQ ID NO: 1062)	EcoRI
	Reverse	AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG (SEQ ID NO: 1063)	PstI

<b>ORF130</b>	Forward	AAAGAATTC-GCAGTACTTGCCATTCTCGGTGCG (SEQ ID NO: 1064)	EcoRI
	Reverse	AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT (SEQ ID NO: 1065)	PstI
<b>ORF 131</b>	Forward	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT (SEQ ID NO: 1066)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAGCGGACGCGTTC (SEQ ID NO: 1067)	XhoI
<b>ORF 132</b>	Forward	GCGGATCCCATATG-AAAGAAGCGGGGTTTG (SEQ ID NO: 1068)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CCAATCTGCCAGCCGT (SEQ ID NO: 1069)	XhoI
<b>ORF 133</b>	Forward	CGCGGATCCCATATG-GAAGATGCAGGGCGCG (SEQ ID NO: 1070)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACTTGTAGCTCATCGT (SEQ ID NO: 1071)	XhoI
<b>ORF 134</b>	Forward	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG (SEQ ID NO: 1072)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCCTGTGCCAATGCG (SEQ ID NO: 1073)	XhoI
<b>ORF 135</b>	Forward	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT (SEQ ID NO: 1074)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAATACCGCTGAGGATG (SEQ ID NO: 1075)	XhoI
<b>ORF 136</b>	Forward	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC (SEQ ID NO: 1076)	BamHI-NheI
	Reverse	CCCGCTCGAG-TTCCGAATATTTGGAACTTTT (SEQ ID NO: 1077)	XhoI
<b>ORF 137</b>	Forward	CGCGGATCCCATATG-GGCACGGCGGGAAATA (SEQ ID NO: 1078)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATAACGGTATGCCGCC (SEQ ID NO: 1079)	XhoI
<b>ORF 138</b>	Forward	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC (SEQ ID NO: 1080)	BamHI-NdeI
	Reverse	CCCGCTCGAG-CGGCGTTTTATAGCGG (SEQ ID NO: 1081)	XhoI
<b>ORF 139</b>	Forward	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG (SEQ ID NO: 1082)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAACGTTTCCGTGCGTTT (SEQ ID NO: 1083)	XhoI
<b>ORF 140</b>	Forward	GCGGATCCCATATG-TTGCCACAGGCAGC (SEQ ID NO: 1084)	BamHI-NdeI

ORF 141	Reverse	CCCGCTCGAG-GACGATGGCAAACAGC (SEQ ID NO: 1085)	XhoI
	Forward	GCGGATCCCATATG-CCGTCTGAAGCAGTCT (SEQ ID NO: 1086)	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCTGTTGTTTTTAAATATT (SEQ ID NO: 1087)	XhoI
ORF 142	Forward	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG (SEQ ID NO: 1088)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGTATAGCCTACCT (SEQ ID NO: 1089)	XhoI
ORF 143	Forward	GCGGATCCCATATG-GATACCGCTTTGAACCT (SEQ ID NO: 1090)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AATGGCTTCCGCAATATG (SEQ ID NO: 1091)	XhoI
ORF 144	Forward	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC (SEQ ID NO: 1092)	BamHI-NdeI
	Reverse	CCCGCTCGAG-AGATTGTTGTTGTTTTTTCG (SEQ ID NO: 1093)	XhoI
ORF 147	Forward	GCGGATCCCATATG-TCTGTCTTTCAAACGGC (SEQ ID NO: 1094)	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTTTTGCAAGACAG (SEQ ID NO: 1095)	XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH<sub>4</sub>OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD<sub>260</sub> was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

### C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including

1.5mM MgCl<sub>2</sub>), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

5 After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	<b>Denaturation</b>	<b>Hybridisation</b>	<b>Elongation</b>
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds 95°C	30 seconds 65-70°C	30-60 seconds 72°C

10

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

15 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

**D) Digestion of PCR fragments**

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs ) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

**E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)**

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the vector pTRC99 (Pharmacia).



**F) Cloning**

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml ).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g ) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SalI* or, for ORF 122, *SalI-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

**G) Expression**

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD<sub>600</sub> ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

**H) GST-fusion proteins large-scale purification.**

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD<sub>280</sub> of 0.02-0.06. The GST-fusion

protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD<sub>280</sub> was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

#### **I) His-fusion solubility analysis (ORFs 111-129)**

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub> PO<sub>4</sub>] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidium-HCl for their solubilization.

#### **J) His-fusion large-scale purification.**

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni<sup>2+</sup>-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD<sub>280</sub> of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D<sub>280</sub> of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D<sub>280</sub> was 0.1. 21µl of each fraction were loaded on a 12% SDS gel.

#### **K) His-fusion proteins renaturation**

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

**L) His-fusion large-scale purification (ORFs 111-129)**

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

**10 M) Mice immunisations**

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)<sub>3</sub> as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than AL(OH)<sub>3</sub>, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

**N) ELISA assay (sera analysis)**

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing

buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN<sub>3</sub> in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were  
5 washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H<sub>2</sub>O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H<sub>2</sub>SO<sub>4</sub> was added to each well and  
10 OD<sub>490</sub> was followed. The ELISA was considered positive when OD<sub>490</sub> was 2.5 times the respective pre-immune sera.

#### **O) FACScan bacteria Binding Assay procedure.**

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and  
15 inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN<sub>3</sub>) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking  
20 buffer to reach OD<sub>620</sub> of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)<sub>2</sub> goat anti-mouse, diluted 1:100, was added to each well and plates  
25 incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H

threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

#### **P) OMV preparations**

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml  
5 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria  
disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were  
removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered  
by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from  
the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and  
10 incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10  
minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75  
minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-  
HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a  
standard.

#### **15 Q) Whole Extracts preparation**

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of  
20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

#### **R) Western blotting**

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg)  
20 derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose  
membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 %  
Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at  
4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was  
washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated  
25 for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed  
twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-  
mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the  
Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

**S) Bactericidal assay**

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD<sub>620</sub> was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD<sub>620</sub> of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II gives a summary of the cloning, expression and purification results.

**TABLE II – Summary of cloning, expression and purification**

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion



orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	
orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	

orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

**Example 1**

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 1):

```

5      1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
      51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101  A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151  TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201  GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251  GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
10  301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401  TCGAAGCGGT CAGATGGTTT CCGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451  GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANC GC GTGCGCCA
501  AGACCG...
```

15 This corresponds to the amino acid sequence (SEQ ID NO: 2; ORF37):

```

      1  MKQTVXMLAA ALIALGLNRP.VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
     51  YXQRTRVRRD DAEAVRWYRQ PAEQGLAQAO YNLGWMYANG RXVRQDDTEA
```

101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ  
 151 AQNNLGVMYA ERXVRQD...

Further work revealed the complete nucleotide sequence (SEQ ID NO: 3):

5           1 ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT  
           51 GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTCGG GAAAACTTGC  
         101 AGGCGGCAGC ACAGGGAAT GCAGCAGCCC AATACAATTT GGGCGCAATG  
         151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG  
         201 GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCA TACAATTTGG  
 10          251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG  
         301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA  
         351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGC GC CAAGACGATG  
         401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA  
         451 GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA  
 15          501 AGACCGCGCC CTTGCACAA AATGGTTTGG CAAGGCTTGT CAAAACGGAG  
         551 ACCAAGACGG CTGCGACAA TACCAACGCC TGAAGCGGG TTATTGA

This corresponds to the amino acid sequence (SEQ ID NO: 4; ORF37-1):

20           1 MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM  
           51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAG YNLGWMYANG RGVRRDDTEA  
         101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ  
         151 AQNNLGVMYA ERRGVRRQDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY\*

Further work identified the corresponding gene in strain A of *N.meningitidis* (SEQ ID NO: 5):

25           1 ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT  
           51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTCGG GAAAACTTGC  
         101 AGGCGGCAGC ACAGGGAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG  
         151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG  
         201 GCTTGGCAAG GCTTGTCAA ACAGGATACCA AGACAGCTGC GACAATGACC  
 30          251 AACGCCTGAA AGCGGGTTAT TGA

This encodes a protein having amino acid sequence (SEQ ID NO: 6; ORF37a):

1 MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGV  
 51 YAERRGVRRD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY \*

35 The originally-identified partial strain B sequence (ORF37) (SEQ ID NO: 2) shows 68.0% identity over a 75aa overlap with ORF37a (SEQ ID NO: 6):

		10	20	30	40	50	60
orf37.pep		MKQTVXMLAAALIALGLNRPVWXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVR RD					
		:					
40	orf37a	MKQTVKWLAAALIALGLNQAVWADDVSDFR ENLQAAAQGNAAAQNNLGVMYAERRGVRRD					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf37.pep		DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG					
		:   : :					
45	orf37a	RALAQEWLGKACQNGYQDSCDNDQRLKAGYX					
		70	80	90			

Further work identified the corresponding gene in *N.gonorrhoeae* (SEQ ID NO: 7):

	1	ATGAAACAGA	CAGTCAAATG	GCTTGCCGCC	GCCCTGATTG	CCTTGGGCTT
5	51	GAACCAAGCG	GTGTGGGCGG	GTGACGTATC	GGATTTTCGG	GAAACTTGC
	101	AGgcggaGA	ACaggGAAAT	GCAGCAGCCC	AATTCaATTT	GGGCGTGATG
	151	TATGAAAATG	GACAAGGAGT	TCGTCAAGAT	TATGTACAGG	CAGTGCAGTG
	201	GTATCGCAAG	GCTTCAGAAC	AAGGGGATGC	CCAAGCCCAA	TACAATTTGG
	251	GCTTGATGTA	TTACGATGGA	CGCGGCGTGC	GCCAAGACCT	TGCCTCGCT
10	301	CAACAAATGGC	TTGGCAAGGC	TTGTCAAAAC	GGAGACCAAA	ACAGCTGCGA
	351	CAATGACCAA	CGCCTGAAGG	CGGGTTATTA	A	

This encodes a protein having amino acid sequence (SEQ ID NO: 8; ORF37ng):

15

1	MKQTVKWLAA	ALIALGLNQA	VWAGDVSDFR	ENLQAAEQGN	AAAQFNLGVM
51	YENGQGVQRD	YVQAVQWYRK	ASEQGDAQAQ	YNLGLMYDYG	RGVRQDLALA
101	QQWLKGACQN	GDQNSCDNDQ	RLKAGY*		

The originally-identified partial strain B sequence (ORF37) (SEQ ID NO: 2) shows 64.9% identity over a 111aa overlap with ORF37ng (SEQ ID NO: 8):

20	orf37.pep	MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAAQGNAAAQYNLGAMYXQRTVRRD	60
		:                        :   :    :   :	
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAAQFNLGVMYENGQGVQRD	60
	orf37.pep	DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG	120
		::  :   : :                  :         :   :   :	
	orf37ng	YVQAVQWYRKASEQGDAAQAYNLGLMYYDGRGVQRDLALAAQQLGKACQNGDQNSCNDNQ	120
25	orf37.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERXVRQD	168
	orf37ng	RLKAGY	126

30 The complete strain B sequence (ORF37-1) (SEQ ID NO: 4) and ORF37ng (SEQ ID NO: 8) show 51.5% identity in 198 aa overlap:

```

35      10      20      30      40      50      60
orf37-1.pep  MKQTVKWLAAALIALGLNRAVWADDVSDFRENLQAAAQGNAAAQYNLGAMYYKGRGVRRD
            |||||:|||||:|||||:|||||:|||||:|:|:|:|:|
orf37ng      MKQTVKWLAAALIALGLNQAVWAGDVSDFFRENLQAAEQGNAAAQFNLGVMYENGQGVRQD
            10      20      30      40      50      60

            70      80      90      100     110     120
orf37-1.pep  DAEAVRWYRQAEEQGLAQAYNLGWMYANGRGVRQDDTEAVRWYRQAAAQGVVQAQYNLG
            ::||:||||:||||| |||:|||||
orf37ng      YVQAVQWYRKASEQGDAQAYNLGLMYDGRGVRRD-----
            70      80      90

            130     140     150     160     170     180
orf37-1.pep  VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERRGVRRDRLAQEWFGKAC
            ||||:||||
orf37ng      -----LALAQQWLKAK
                        100

```

```

                    190      199
orf37-1.pep      QNGDQDGCNDNDQRLKAGYX
                  |||||:::|||||
orf37ng          QNGDQNSCDNDQRLKAGYX
                  110      120

```

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 10 ORF37-1 (SEQ ID NO: 4) (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 15 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 (SEQ ID NO: 4) is a surface-exposed protein, and that it is a useful immunogen.

Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1 (SEQ ID NO: 4).

## Example 2

- 20 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 9):

```

TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA
TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
ACGTTTCCGC GCAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC
25 GGGCTGCAGC AGGGCGGCGA CACGGAAC CTTGCTGCCG GCGACACCAT
CTCCGTAACC AGTTCTGCAA TGGTCTGGA AAACCTTATC GGCAAATTCA
TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GGAAAAAGCC
GCCGAATAA

```

- 30 This corresponds to the amino acid sequence (SEQ ID NO: 10):

```

1  FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD
51 VSAQILTSLG LGEQYIGLQQ GGDENLAAG DTISVTSSAM VLENLIGKFM
101 TSFAEKNADG GNAEKAAE*

```

- 35 Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029 (SEQ ID NO: 1105))

SEQ ID NO: 9 and ybrd.haein (SEQ ID NO: 1105) show 48.4% aa identity in 122 aa overlap:

5		20	30	40	50	60	70
	ybrd.h	LGIGALVFLGLRVANVQGFAETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE					
	N.m				FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP		
					10	20	30
10		80	90	100	110	120	130
	ybrd.h	KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT					
	N.m	KSYQARVRDLDDGKY-QFSSDVSAQILTSGLLGEQYIGLQQG--GD TENLAAGDTISVT					
		40	50	60	70	80	
15		140	150	160			
	ybrd.h	TSAMVLEDLIGQFL--YGSKSDGNEKSESTEQ					
	N.m	SSAMVLENLIGKFMTSFAEKNADGGNAEKAEX					
		90	100	110	120		

Homology with a predicted ORF from *N.gonorrhoeae*

20 SEQ ID NO: 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae* (SEQ ID NO: 1106 yrbx):

25		20	30	40	50	60	70
	ybrd	GAAAVAFLAFRVAGGAFFGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP					
	N.m				FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP		
					10	20	30
30		80	90	100	110	120	130
	ybrd	KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM					
	N.m	KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM					
		40	50	60	70	80	90
35		140	150	160			
	ybrd	VLENLIGKFMTSFAEKNADGGNAEKAEX					
	N.m	VLENLIGKFMTSFAEKNADGGNAEKAEX					
		100	110	120			

The complete ybrd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

**Example 3**

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 11):

```

1  . .ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTCGCCCC TCTTCTTCTT
51  TCAGGAACGC CCCGGAAGG ACGGAAAACC TTTTAAAATG GTCAAATTCC
5   101  GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
151  GAACGCCTGA CACCGTTCGG CAAAAAATG CGTGCCGcCA GTwTGGACGA
201  ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
251  CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
10  301  CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG
351  GCGCAACGCg CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
401  TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451  AAAAAAGTAT TAATCAAGGA AGGGATTTC GCACAGGGCG AACA .aCCAT
501  GCCCCCTTTC ACAGGAAAAC GCAAATCGC CGTCGTCGGT GCGGGCGGAC
551  ACGGAAAAGT CGTGGCCGAC CTGCGCCCG CACTCGGCCG GTACAGGAA
15  601  ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
651  CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
701  ACGTCGCCGT CGCCGTCGGC AACAAACGCA TCCGCCGCCA AATCGCCGAA
751  AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
801  GACCGTCTCG CTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
20  851  AAGCGGTCG. .

```

This corresponds to the amino acid sequence (SEQ ID NO: 12; ORF3):

```

1  . .ILIIYLIRKNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
51  ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
25  101  RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTV
151  KKVLIKEGIS AQGEXTMPFF TGKRKLAVVG AGGHGKVVD LAAALGRYRE
201  IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
251  KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV. .

```

30 Further sequence analysis revealed the complete nucleotide sequence (SEQ ID NO: 13):

```

1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
51  ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
101  AGAATCTAGG TTCGCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
35  151  GGAAAACCTT TTAATATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA
201  TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
251  AAAAACTGCG TGCCGCCAGT TTGGACGAAC TGCCTGAATT ATGGAATATC
301  TTAATAAGGCG AGATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
351  TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
401  GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
40  451  GAAAAATTCG CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
501  CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
551  GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
601  AAACTCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
651  TGCCGCCGCA CTCCGCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
45  701  CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
751  GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
801  CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCCGCGCG CTCGGCTTCG
851  CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
901  GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
50  951  CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001  ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTGC
1051  GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG

```

1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG  
 1151 TCGTCGTACG CGACGTTTCA GACGCATGA CCGTCGCGGG CAATCCGGCA  
 1201 AAGCCGCTGC CGCGCAAAAA CCCCGAGACC TCGACAGCAT AA

5 This corresponds to the amino acid sequence (SEQ ID NO: 14; ORF3-1):

1 MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSPV FFFQERPGKD  
 51 GKPFKMKVFR SMRDALDSGD IPLPDGERLT PFGKKLRAAS LDELPELWNI  
 101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD  
 151 EKFACDVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPTGKR  
 201 KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLLL  
 251 ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT  
 301 VGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISPAHLS  
 351 GNTHIGESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA  
 401 KPLPRKNPET STA\*

15 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 (SEQ ID NO: 12) shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) (SEQ ID NO: 16) from strain A of *N. meningitidis*:

20 orf3.pep 10 20 30  
 ILIYLIRKNLGSPVFFFQERPGKD GKPFKMKVFR  
 orf3a MSKFFKRLFDIVASASGLIFLSPVFLILY LIRKNLGSPVFFFQERPGKD GKPFKMKVFR  
 10 20 30 40 50 60

25 orf3.pep 40 50 60 70 80 90  
 SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGP RPLLMQYLPL  
 orf3a SMHDALDSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGEMSLVGP RPLLMQYLPL  
 70 80 90 100 110 120

30 orf3.pep 100 110 120 130 140 150  
 YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFS LCLDIKILLTVKKVL  
 orf3a YDNFQNRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFS LCLDIKILLTVKKVL  
 130 140 150 160 170 180

35 orf3.pep 160 170 180 190 200 210  
 IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG  
 orf3a IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAEALAAALGTGEIVFLDDRQGSVNG  
 190 200 210 220 230 240

40 orf3.pep 220 230 240 250 260 270  
 FSVIGTLLLENSLSPEQYDVAVAVGNNRI RRQIAEKAAAALGFALPVLVHPDATVSPSAT  
 orf3a FSVIGTLLLENSLSPEQFDIAVAVGNNRI RRQIAEKAAAALGFALPVLVHPDSTVSPSAT  
 250 260 270 280 290 300

45 280



```

orf3.pep      VGQGSVMAKAV
               ||||:|||||
orf3a         VGQGGVMAKAVVQADSVLKDGIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEE
               310          320          330          340          350          360

```

5 The complete length ORF3a nucleotide sequence (SEQ ID NO: 15) is:

	1	ATGAGTAAAT	TCTTCAAACG	CCTGTTTGAC	ATTGTTGCCT	CCGCTCGGG
	51	ACTGATTTTC	CTCTCGCCAG	TATTTTGTAT	TTTGATATAC	CTCATCCGCA
	101	AGAATCTGGG	TTCGCCCGTC	TTCTTCTTTC	AGGAAACGCC	CGGAAAGGAC
10	151	GGAAAACCTT	TTAAAAATGGT	CAAATTCGGT	TCCATGCACG	ACGCGCTTGA
	201	TTCAGACGGG	ATTCGTCTGC	CCGACGGAGA	ACGCTGTGAC	CCGTTCCGGC
	251	AAAAACTGCG	TGCCGCCAGT	TTGGACGAAC	TGCCCGAACT	GTGTCAAGCTC
	301	CTCAAAGGCG	ACATGAGCCT	GGTCGGCCCC	CGCCCGCTGC	TGATGCAATA
	351	TCTGCCGCTG	TACGACAACT	TCCAAAACCG	CCGCCACGAA	ATGAAAACGG
15	401	GCATTACGGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
	451	GAACGCTTCG	CATGCGACAT	CTGGTATATC	GACCACTTCA	GCCTGTGCCT
	501	CGACATCAAA	ATCCTACTGC	TGACGGTTAA	AAAAGTATTA	ATCAAAGAAG
	551	GGATTTCCGC	ACAGGGCGAA	GCCACCATGC	CCCCTTTCAC	AGGAAAACGC
	601	AAACTTGCCG	TCGTCTGGTG	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
20	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCG
	701	TCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCGCCGTGC	CCGTCCGGCAA
	801	CAACCGCATC	CGCCGCCAAAT	TCGCCGAAAA	AGCCGCCCGG	CTCGGCTTCC
	851	CCCTGCCCGT	CCTGATTCAT	CCGGACTCGA	CCGTCTCGCC	TTCTGCAACA
25	901	GTCCGACAA	GCGGCGTCGT	TATGGCGAAA	GCCGTCGTAC	AGGCTGACAG
	951	CGTATTGAAA	GACGGCGTAA	TTGTGAACAC	TGCCGCCACC	GTGATCAAG
	1001	ATTGCCTGCT	TGATGCTTTC	GTCCACATCA	GCCCGGGCGC	GCACCTGTCT
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCTGG	ATAGGCACAG	GCGCGTGCAG
	1101	CCGCCAGCAG	ATCCGTATCG	GCAGCCCGCG	AACCAATTGA	GCGGGCGCAG
30	1151	TCGTCTGTGC	CGACGTTTCA	GACGGCATGA	CCGTCCGGGG	CAACCCGGCA
	1201	AAACCATTGG	CAGGCAAAAA	TACCGAGACC	CTGCCGTCTG	AA

This is predicted to encode a protein having amino acid sequence (SEQ ID NO: 16):

	1	MSKFFKRLFD	IVASASGLIF	LSPVFLILIY	LIRKNLGPV	FFQERPGKD
35	51	GKPFKMKVFR	SMHDALDSG	ILLPDGERLT	PFGKKLRAAS	LDELPELWNV
	101	LKGDSL VG	RPLLMQYLPL	YDNFQNRHE	MKPGITGWAQ	VNGRNALSWD
	151	ERFACDIWYI	DHFSLCLDIK	ILLT VKKVL	IKEGISAQGE	ATMPPTGKR
	201	KLAVVGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRVQGSVNG	FPVIGTTL
	251	ENSLSPEQFD	IAVAVGNMRI	RRQIAEKAAA	LGFALPVLH	PDSTVSPSAT
40	301	VGQGGVMAK	AVVQADSVLK	DGVIVNTAAT	VDHDCLLDAF	VHISPGAHL
	351	GNTRIGEESS	IGTGACSRQQ	IRIGSRATIG	AGAVVVRDVS	DGMTVAGNPA
	401	KPLAGKNTET	LRS*			

Two transmembrane domains are underlined.

ORF3-1 (SEQ ID NO: 14) shows 94.6% identity in 410 aa overlap with ORF3a (SEQ ID NO: 16):

45		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILIIYLIRKNLGSPVFFFQERPGKDGGKPFKVMVKFR					
	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILIIYLIRKNLGSPVFFFQERPGKDGGKPFKVMVKFR					
		10	20	30	40	50	60
50		70	80	90	100	110	120
	orf3a.pep	SMHDA LDS DGI LL PDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLL MQYLPL					

	orf3-1	SMRDALDSGIPLPDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRLPLMQYLP 70 80 90 100 110 120
5	orf3a.pep	YDNFQNRHMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSCLDIKILLTVKKVL 130 140 150 160 170 180
	orf3-1	YDNFQNRHMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLDIKILLTVKKVL 130 140 150 160 170 180
10	orf3a.pep	IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRVQGSVNG 190 200 210 220 230 240
	orf3-1	IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAELAAALGRYREIVFLDDRAQGSVNG 190 200 210 220 230 240
15	orf3a.pep	FPVIGTTLLLENSLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLHPDSTVSPSAT 250 260 270 280 290 300
	orf3-1	FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT 250 260 270 280 290 300
20	orf3a.pep	VGQGGVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGESW 310 320 330 340 350 360
	orf3-1	VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGESW 310 320 330 340 350 360
25	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLAGKNTETLR SX 370 380 390 400 410
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX 370 380 390 400 410

30 Homology with hypothetical protein encoded by yvfc gene (accession Z71928) (SEQ ID NO: 1108) of *B. subtilis*

ORF3 (SEQ ID NO: 12) and YVFC proteins (SEQ ID NO: 1108) show 55% aa identity in 170 aa overlap (BLASTp):

35	ORF3	3	IYLIRKNLGS PVFFFQERPGKDGKPFKMKVFRSMRDGLYSDGIPLPDGERLTPFGKKLRA	62
	yvfc	27	I + + R + GSPVFF Q RPG GKPF + KFR + M D S G LPD RLT G + + R IAVRLKIGSPVFFKQVRPGLHGKPF TLYKFR TMTDERDSKGNLLPDEVRLTKTGRLIRK	86
	ORF3	63	ASXDELPELWNILKGEMSLVGPRLPLMQYLPYDNFQNRHMKPGITGWAQVNGRNALS	122
	yvfc	87	S DELP + L N + LKG + + SLVGPRLPLM YLPLY Q RRHE + KPGITGWAQ + NGRNA + S LSIDELPQLLNVLKGDLSLVGPRLPLMDYLPLYTEKQARRHEVKPGITGWAQINGRNAIS	146
40	ORF3	123	WDEKFACDVWYIDHFSCLDXXXXXXXXXXXXXXXXXEGISAQGEEXTMPPTG	172
	yvfc	147	W + + KF DVWY + D + + S LD EGI T FTG WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEGIQQTNHVTAERFTG	196

Homology with a predicted ORF from *N. gonorrhoeae*

ORF3 (SEQ ID NO: 12) shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) (SEQ ID NO: 18) from *N. gonorrhoeae*:

	orf3	ILIIYLIRKKNLGSPVFFFQERPQKDGKPFKMVKFR	34
		:             ::	
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKKNKGSPVFFIRERPGKDGKPFKMVKFR	60
	orf3	SMRDGLYS DGIPDPGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLLPL	94
		:             :           :	
	orf3ng	SMRDALDS DGIPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLLPL	120
	orf3	YDNFQNR RHMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLCDIKILLTVKKVL	154
10		::	
	orf3ng	YNKFQNR RHMKPGITGWAQVNGRNALSWDEKFS CDVWYTDNFSFWLDMKILFLT VKKVL	180
	orf3	IKEGISAQGE XTMPPTGKRKLAVVGAGGHGKVADLAAALGRYREIVFLDDRAQGSVNG	214
		:       :	
	orf3ng	IKEGISAQGEATMP PFAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG	240
15	orf3	FSVIGTTLLLENSLSPEQYDVA VAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
		: ::	
	orf3ng	FPVIGTTLLLENSLSPEQFDITVA VGNRRIRRQITENAAALGFKLPVLIHPDATVSPSAI	300
	orf3	VGQGSVVMAKAV	286
		:	
20	orf3ng	IGQGSVVMAKAVVQAGSVLKDGVI VNTAATVDHDCLLDAFVHISPGAHL SGNTRIGEE SR	360

The complete length ORF3ng nucleotide sequence (SEQ ID NO: 17) is:

	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTGCGCCG	TGTTTGTGGT	TTTAATATAC	CTCATCCGCA
25	101	AAAACCTAGG	TTCGCCCCGTC	TTCTTCattc	GGGAACGCCc	cgGAAAGGAc
	151	ggaaaacCTT	TTAAATGGT	CAAATTCGT	TCCatgcgcg	acgcgcttGA
	201	TTCAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCGGCA
	251	AAAAATTACG	CGCCACCAGT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
	301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
30	351	TCTGCCGCTT	TACAACAAAT	TTCAAACCG	CCGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACCGCT	TTCGTGGGAC
	451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
	501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
	551	GCATTTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
35	601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
	701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCACCGTCG	CCGTCCGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
40	851	AAC TGCCCGT	TCTGATT CAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
	901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCTGAC	AGGCCGCGAG
	951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
	1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCTG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
45	1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgccgGT	GCAGGGgcGG
	1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTGCGGGG	CAACCCGGCA
	1201	AAGCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence (SEQ ID NO: 18):

```

1  MSKAVKRLFD IIASASGLIV LSPVFLVLIY LIRKNLGSPV FFIRERPGKD
51  GKPFKMVKFR SMRDALDSDG IPLPDSERLT DFGKKLRATS LDELPELWNV
101 LKGEMSLVGP RPLLMQYLPL YNKFNRRHE MKPGITGWAQ VNGRNALSWD
151 EKFSQCDVWYT DNFSFWLDMK ILFLTVMKVL IKEGISAQGE ATMPPFAGNR
201 KLAIVAGAGH GKVVAELAAA LGTYGEIVFL DDRTQGSVNG FPGVIGTLLL
251 ENSLSPEQFD ITVAVGNNRI RRQITENAAA LGFKLPVLIH PDATVSPSAI
301 IGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLLDFA VHISPGAHLS
351 GNTRIGESR IGTGACSRQQ TTVGSGVTAG AGAVIVCDIP DGMTVAGNPA
401 KPLTGKNPKT GTA*

```

This protein shows 86.9% identity in 413 aa overlap with ORF3-1 (SEQ ID NO: 14):

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLIILIYLRKNLGSPVFFQERPGKDGKPFKMVKFR					
		:             :             :					
15	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKNLGSPVFFIRERPGKDGKPFKMVKFR					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf3-1.pep	SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNLKGMESLVGPRPLLMQYLPL					
		:             :             :					
20	orf3ng	SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNLKGMESLVGPRPLLMQYLPL					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf3-1.pep	YDNFQNRHMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLDIKILLTVKKVL					
		:             :             :             :					
25	orf3ng	YNKFQNRHMKPGITGWAQVNGRNALSWDEKFSQCDVWYTDNFSFWLDMKILFLTVMKVL					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf3-1.pep	IKEGISAQGEATMPPFAGNRKLAIVAGAGHGKVVADLAAALGRYREIVFLDDRTQGSVNG					
		:             :             :					
30	orf3ng	IKEGISAQGEATMPPFAGNRKLAIVAGAGHGKVVADLAAALGRYREIVFLDDRTQGSVNG					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf3-1.pep	FSVIGTLLLENSLSPEQYDVAVAGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
		:             :             :					
35	orf3ng	FPVIGTLLLENSLSPEQFDITVAVAGNNRIRRQITENAAALGFALPVLVHPDATVSPSAI					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf3-1.pep	VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGESW					
		:             :             :             :					
40	orf3ng	IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLDFAVHISPGAHLSGNTRIGESR					
		310	320	330	340	350	360
		370	380	390	400	410	
	orf3-1.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
		:             :             :					
45	orf3ng	IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTA					
		370	380	390	400	410	

In addition, ORF3ng (SEQ ID NO: 18) shows significant homology with a hypothetical protein (SEQ ID NO: 1110) from *B.subtilis*:

gnl|PID|e238668 (Z71928) hypothetical protein [Bacillus subtilis]  
 )gi|1945702|gnl|PID|e313004 (Z94043) hypothetical protein [Bacillus subtilis]  
 )gi|2635938|gnl|PID|e1186113 (Z99121) similar to capsular polysaccharide  
 biosynthesis [Bacillus subtilis]Length = 202

Score = 235 bits (594), Expect = 3e-61

Identities = 114/195 (58%), Positives = 142/195 (72%)

Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSPVFFIRERPGKDGKPFKPMVKFRSMRD 64  
 +KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D

Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILFTIAVVRLKIGSPVFFKQVRPGLHGKPFPLYKFRMTMD 62

Query: 65 ALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPLYNKF 124  
 DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLM YLPLY +

Sbjct: 63 ERDSKGNLLPDEVRLTKTGRLLRKLSIDELPQLLNVLKGDLSLVGPRPLLM DYLYTEK 122

Query: 125 QNRRHEMKPGITGWAQVNGRNALSWDEKFS CDVWYTDNFSFWLDMKILFLTVKKVLIKEG 184  
 Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG

Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEG 182

Query: 185 ISAQGEATMPPFAGN 199

I T F G+

Sbjct: 183 IQQTNHVTAEFTGS 197

The hypothetical product of *yvfc* gene shows similarity to EXOY of *R.meliloti*, an  
 exopolysaccharide production protein. Based on this and on the two predicted transmembrane  
 regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their  
 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 4

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 19):

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GGCGGCACAT CCGGCTTGGT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
101 TTAGCAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
401 CAGTTTAG

```

This corresponds to amino acid sequence (SEQ ID NO: 20; ORF5):

```

1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSA RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REVS....XR RFCTV*

```

Further sequence analysis revealed the complete DNA sequence to be (SEQ ID NO: 21):

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCG

```

5 51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC  
 101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA  
 151 AAAGTCCTCG ATTTTTCGA TTTGGAAGTG CGCGACGCGA TGATTACGCG  
 201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG  
 251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGCGAAGAC  
 301 AAAGACGAAG TTTTGGGCGAT TTTGCACGCC AAAGACCTGC TCAAATATAT  
 351 GTTTAAACCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT  
 401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA  
 10 451 CAGCGCAACC ATATGCGGAT TGTATCGAC GAATACGGCG GCACATCCGG  
 501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG  
 551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC  
 601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT  
 651 CTTCCGCACG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC  
 701 ATTCAAGAGT TGGGACATCT GCCCGTGC GC GCGAAAAAG TCCTTATCGG  
 15 751 CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC  
 801 TGATGCGGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA  
 851 TGACGGTACG GCGGTTTTCT GTTCAATCC GCCCATCCG CCAAACATAA

This corresponds to amino acid sequence (SEQ ID NO: 22; ORF5-1):

20 1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLRLLE  
 51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED  
 101 KDEVLGILHA KDLLKYMFPN EQFHLKSILR PAVFVPEGKS LTALLKEFRE  
 151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS  
 201 ERWRIHAATE IEDINTFFGT EYSSEADTI RPHGSRVGTS ARARRKSPYR  
 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT\*

Further work identified the corresponding gene in strain A of *N.meningitidis* (SEQ ID NO: 23):

30 1 ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCC  
 51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC  
 101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA  
 151 AAAGTCCTCG ATTTTCTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG  
 201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG  
 251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGTGAAGAC  
 301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT  
 35 351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT  
 401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA  
 451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG  
 501 TTTGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGGC GACATCGAAG  
 551 ATGAGTTTGA CGAAGACGAA AGCGCGGACA ACATCCACGC CGTTTCCGCC  
 40 601 GAACGCTGGC GCATCCACGC GGCTACCGAA ATCGAAGACA TCAACGCCTT  
 651 TTTCCGCACG GAATACAGCA GCGAAGAAGC CGACACCATC GGCGGCCNTG  
 701 GTCATTCAAG AAT'TGGNACA CCTGCCCGTG CGCGGCGAAA AAGTCNTTAT  
 751 CGGCGNNTTG CANTTACNG TCGCCNGCGC NGACAACCGC CGCCTGCATA  
 801 CGCTGATGGC GACCCGCGTG AAGTAAGCTC CGCCGTTTCT GTACAGTTTA  
 45 851 GGATGACGGT ACGGGCGTTT TCTGTTTCAA TCCGCCCCAT CCGCCANACA  
 901 TAA

This encodes a protein having amino acid sequence (SEQ ID NO: 24; ORF5a):

50 1 MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLRLLE  
 51 KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED  
 101 KDEVLGILHA KDLLKYMFPN EQFHLKSILR PAVFVPEGKS LTALLKEFRE  
 151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADNIHAVSA  
 201 ERWRIHAATE IEDINAFPGT EYSSEADTI GGXGHSIGT PARARRKSXY  
 251 RRXAXHXRXR XQPPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT

301 \*

The originally-identified partial strain B sequence (ORF5) (SEQ ID NO: 20) shows 54.7% identity over a 124aa overlap with ORF5a (SEQ ID NO: 24):

```

5      10      20      30
orf5.pep      NHMAIVIDEYGGTSGGLVTFEDIIEQIVGEI
orf5a      FHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVGDI
           130      140      150      160      170      180

10     40     50     60     70     80     90
orf5.pep      EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA
orf5a      EDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADTIGGXGHSIGGTPA
           190      200      210      220      230      240

15     100     110     120     130
orf5.pep      RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSSXXXXRRFCTV
orf5a      RARRKSXYRRXAXHXRXRQPPPAYADGDPREVSSAVSVQFRMTVRAFVSVSIRPIRXTX
           250      260      270      280      290      300

```

The complete strain B sequence (ORF5-1) (SEQ ID NO: 22) and ORF5a (SEQ ID NO: 24) show 92.7% identity in 300 aa overlap:

```

25     10     20     30     40     50     60
orf5a.pep      MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDLFDSDLEV
orf5-1      MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLLRLEKVLDLFDSDLEV
           10     20     30     40     50     60

30     70     80     90     100     110     120
orf5a.pep      RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN
orf5-1      RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFPN
           70     80     90     100     110     120

35     130     140     150     160     170     180
orf5a.pep      EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG
orf5-1      EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG
           130     140     150     160     170     180

40     190     200     210     220     230     240
orf5a.pep      DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADTIGGXGHSIGG
orf5-1      EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT
           190     200     210     220     230

45     250     260     270     280     290     300
orf5a.pep      PARARRKSXYRRXAXHXRXRQPPPAYADGDPREVSSAVSVQFRMTVRAFVSVSIRPIRXT
orf5-1      SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFVSVSIRPIRQT
           240     250     260     270     280     290

```

Further work identified the a partial DNA sequence in *N.gonorrhoeae* (SEQ ID NO: 25) which encodes a protein having amino acid sequence (SEQ ID NO: 26; ORF5ng):

```

1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFPGT EYGSEEDTI RRLGHSIGIT PARARRKSPY
251 RRFVHRRPR RQPPAHADG DPREVSRACP HRRFCTV*

```

10 Further analysis revealed the complete gonococcal nucleotide sequence (SEQ ID NO: 27) to be:

```

1 ATGGACGGCG CACAACCGAA AACAAATTTT TTTGAACGCC TGATTGCCCC
51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA ACAGGAAGTT TTTGATGCCG ACACACTGAC CCGGCTGGAA
151 AAAGTATTGG ACTTTGCCGA GCTGGAAGTG CGCGATGCGA TGATTACGCG
201 CAGCCGCATG AACGTATTGA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTCAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTGAAATC CGTCTTGCGC CCTGCCGTTT
401 TCGTGCCCGA AGGCAAATCT TTGACCGCCC TTTTAAAAGA GTTCCGCGAA
451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
511 TTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGT GACATCGAAG
551 ACGAGTTTGA CGAAGACGAA AGCGccgacg acatCCACTC cgTTTccgCC
601 GAACGCTGGC GCATCCacgc ggctaCCGAA ATCGAAGaca TCAACGCCTT
651 TTTCGGTACG GAatacggca gcgaagaagc cgacaccatc cggcggtTG
701 GTCATTACAG AATTGGGACA CCTGCCCGTG CGCGGCGAAA AAGTCCTTat
751 cggcgGTTTG Cagttcacgc tCGCCCGCGC CGACAACCGC CGCCTGCACA
801 CGCTGATGGC GACCCGCGTG AAGTAAGCAG AGCCTGCCcg AccgcggtT
851 CTGCacAGTT TAGGatGACG gtaCGGTCGT TTTCTGTTC AATCCGCCCC
901 ATCCGCCAAA CATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 28; ORF5ng-1):

```

1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFPGT EYGSEEDTI RRLGHSIGIT PARARRKSPY
251 RRFVHRRPR RQPPAHADG DPREVSRACP TAVSAQFRMT VRSFVSIRP
301 IRQT*

```

40 The originally-identified partial strain B sequence (ORF5) (SEQ ID NO: 20) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng) (SEQ ID NO: 26):

```

orf5                               NHMAIVIDEYGGTSGGLVTFEDIIEQIVGEI 30
                                |||:|||||:|||||:|||||:|
orf5ng    FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVGDI 182
45 orf5    EDEFDEDDSDADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA 90
    |||:||||:||||:||||:|:|||||:|||||:|||||:|||||:|
orf5ng    EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFPGTEYGSEEDTIRRLGHSIGITPA 242
orf5    RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSX----RRFCTV 131

```



orf5ng                    |||||:|||||                    |||||  
                          RARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSACPHRRFCTV    287

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) (SEQ ID NO: 22 & SEQ

5 ID NO: 28) show 92.4% identity in 304 aa overlap:

		10	20	30	40	50	60
	orf5ng-1.pep	MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFAELEV					
10	orf5-1	MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFSDLEV					
		10	20	30	40	50	60
	orf5ng-1.pep	RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP					
15	orf5-1	RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP					
		70	80	90	100	110	120
	orf5ng-1.pep	EQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG					
20	orf5-1	EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIVG					
		130	140	150	160	170	180
	orf5ng-1.pep	DIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSGIGT					
25	orf5-1	EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT					
		190	200	210	220	230	240
	orf5ng-1.pep	PARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSACPTAVSAQFRMTVRSFSVSIRP					
30	orf5-1	SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVS---TAVSAQFRMTVRAFSVSIRP					
		240	250	260	270	280	290
	orf5ng-1.pep	IRQTX					
35	orf5-1	IRQTX					
		300					

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

Homology with hemolysin homolog TlyC (accession U32716) (SEQ ID NO: 1111) of H.influenzae

40 ORF5 (SEQ ID NO: 20) and TlyC proteins (SEQ ID NO: 1111) show 58% aa identity in 77 aa overlap (BLASTp).

ORF5 2 HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDTWRIHAATEIED 61  
HMAIV+DE+G SGLVT EDI+EQIVG+IEDEFDE++ AD I +S T+ + A T+I+D  
TlyC 166 HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

ORF5 62 INTFFGTEYSIEEADTI 78  
N F T++ EE DTI  
TlyC 225 FNAQFNTDFDDEEVDTI 241

ORF5ng-1 (SEQ ID NO: 28) also shows significant homology with TlyC (SEQ ID NO: 1111):

SCORES Initl: 301 Initn: 419 Opt: 668  
Smith-Waterman score: 668; 45.9% identity in 242 aa overlap

		10	20	30	40	50
orf5ng-1.pep		MDGAQPKTNFFERLIARLAR-EPDSAEDVLNLLRQAHEQEVFDADTLTRLEK				
tlyc_haein	MNDEQQNSNQSENTKKPFFQSLFGRFFQGELKNREELVEVIRDSEQNLDLIDQNTREMI					
		10	20	30	40	50
orf5ng-1.pep		60	70	80	90	100
	VLDFAELEVRDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGE--DKDEVLGILH					
tlyc_haein	VMEIAELRVRDIMIPRSQIIFIEDQQDLNTCLNTIIESAHSRFPVIADADDRDNVIGILH					
		70	80	90	100	110
orf5ng-1.pep		110	120	130	140	150
	AKDLLKYM-F-NPEQFHLKSVLRPAVFPVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGL					
tlyc_haein	AKDLLKFLREDAEVFDLSSLLRPVVIVPESKRVRMLKDFRSERFHMAIVVDEFGAVSGL					
		130	140	150	160	170
orf5ng-1.pep		170	180	190	200	210
	VTFEDIIEQIVGDIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEED					
tlyc_haein	VTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEVD					
		190	200	210	220	230
orf5ng-1.pep		230	240	250	260	270
	TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSACPTAVSAQF					
tlyc_haein	TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTPDEHLAEMNNVDEKSE					
		240	250	260	270	280

### Homology with a hypothetical secreted protein from *E.coli*:

ORF5a (SEQ ID NO: 24) shows homology to a hypothetical secreted protein (SEQ ID NO: 1112) from *E.coli*:

sp|P77392|YBEX\_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION  
)gi|1778577 (U82598) similar to H. influenzae [Escherichia coli] )gi|1786879  
(AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an  
approx. 440 aa protein YTFL\_HAEIN SW: P44717 [Escherichia coli] Length = 292

Score = 212 bits (533), Expect = 3e-54  
 Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

Query: 2 DGAQPKTNFXXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60  
 D K F L+++L EP + +++L L+R + + ++ D DT LE V+D +D V  
 5 Sbjet: 10 DTISNKKGFFSLLLSQLFHPGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

Query: 61 RDMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119  
 RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +  
 Sbjet: 70 RDIMIPRSQMITLKRNTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIV 179  
 E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV  
 10 Sbjet: 130 AEAFSMDKVLRLQAVVPESKRVDRLKEFRSQRYHMAIVIDEFGGVSGLVTTIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFPGTEYSSEEADT 229  
 G+IEDE+DE++ D +S W + A IED N FGT +S EE DT  
 15 Sbjet: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVDT 238

Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

ORF5-1 (SEQ ID NO: 22) (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 (SEQ ID NO: 22) is a surface-exposed protein, and that it is a useful immunogen.

### Example 5

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 29):

1 ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTC  
 51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC  
 101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC  
 151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCCCG ACAGCTACGA  
 201 AATCGATGCG GCGGCGAGTG ATTTGCAGAT TTACCAAACC GCCTACAAgG  
 251 GCGATGCAAC GCCGCCTGAA TGA<sup>g</sup>GGCATG GGAAAGCAGG CAGGACGGGC  
 301 TGCCTTATAA AAACCTTAT GAAATGCTGA TTATGGCGAr CCTGGTTCGAA  
 351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT  
 401 CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT  
 451 ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC  
 501 CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC  
 551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence (SEQ ID NO: 30; ORF7):

```

1 MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
51 DAFSGNPEQG FFPDSYEIDA GGS DLQIYQT AYKAMQRR LN EAWESRQDGL
101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFVNRLKIGM RLQTXSVIY
151 GMGAAYKGKI RKADLRDTP YNTYTRGGLP PTPIALP..

```

Further sequence analysis revealed the complete DNA sequence (SEQ ID NO: 31):

```

10 1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCCGC
51 AGCCGTTTTTC GCCGCGCTGC TTTTGTTC TAAGGATAAC GGCAGGGCAT
101 ACCGAATCAA AATTGCCAAA AACCAGGTA TTTCGTCCGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTGA CGGCGGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGACG TACAGATTGC
251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCGAGG
15 301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTTCGCATAT
351 GAGGAAAGTC ATCGACGCAA CGCCGACAT CGGACACGAC ACCAAAGGCT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTCAGCGGC
451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
501 CAGTGATTTC CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCCGCC
20 551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
25 801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
951 GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTGAAA AAATAA

```

30 This corresponds to the amino acid sequence (SEQ ID NO: 32; ORF7-1):

```

1 MLRKLLKWSA VFLTVAADF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLEAWES RQDGLPYKNP
35 201 YEMLIMASLV EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301 FVSKMDGTGL SQFSDHLEH NAAVRKYILK K*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with hypothetical protein encoded by yceg gene (accession P44270) (SEQ ID NO: 1113) of H.influenzae

ORF7 (SEQ ID NO: 30) and yceg proteins (SEQ ID NO: 1113) show 44% aa identity in 192 aa overlap:

```

45 ORF7 1 MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55
+ G+ V+ IEG F RK ++ P + K SNE++ A ++ +
yceg 102 LNSGKEVQFNVKWI EGKTFKDWKDLNAPHLVQTLKDKSNEEIFALLDLPDIGNLELK 161

```

ORF7	56	NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLLNEAWESRDGLPYKNPYEMLIMAXLV	115
		N EG +PD+Y +DL++ + + + M++ LN+AW R + LP NPYEMLI+A +V	
yceg	162	NVEGWLYPDTYNYTPKSTDLELLKRSAERMKKALNKAWNERDEDLPLANPYEMLILASIV	221
ORF7	116	EKETGHEAXXDHVASVFVNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRRDTPYNTYT	175
		EKETG VASVF+NRLK M+LQT +VIYGMG Y G IRK DL TPYNTY	
yceg	222	EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV	281
ORF7	176	RGGLPPTPIALP	187
		GLPPTPIA+P	
yceg	282	IDGLPPTPIAMP	293

The complete length YCEG protein has sequence:

1	KKKFLIAILL	LILILAGVAS	FSYYKMTEFV	KTPVNVQADE	LLTIERGTTS
51	SKLATLFEQE	KLIADGKLLP	YLLKLKPELN	KIKAGTYSLE	NVKTVDLLD
101	LLNSGKEVQF	NVKWIEGKTF	KDWRKDLENA	PHLVQTLKDK	SNEEFALLD
151	LPDIGQNLLEL	KNVEGWLYPD	TYNYPKSTD	LELLKRSAR	MKKALNAWN
201	ERDELDPLAN	PYEMILIAS	VEKETGIANE	RAKAVSVFIN	RLKAKMKLQT
251	DPTVIYGMGE	NYNGNIRKKD	LETKTPYNTY	VIDGLPPTPI	AMPESSSLQA
301	VANPEKTDYF	YFVADSGSGH	KFTRNLNEHN	KAVQEYLRWY	RSQKNAK

### Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF7 (SEQ ID NO: 30) shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) (SEQ ID NO: 34) from strain A of *N. meningitidis*:

```

                                10          20          30
orf7.pep                      MRGGRPDSVTVQIIEGSRFSHMRKVIDATP
                                |||
orf7a      AAYVLGVHNRRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDATP
              70          80          90          100          110          120

              40          50          60          70          80          90
orf7.pep    DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRRLN
              |||
orf7a      DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRRLN
              130          140          150          160          170          180

              100          110          120          130          140          150
orf7.pep    EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFVNRLKIGMRLQTXXSVIY
              |||
orf7a      EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRHVASVFVNRLKIGMRLQTDPSVIY
              190          200          210          220          230          240

              160          170          180
orf7.pep    GMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALP
              |||
orf7a      GMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLYFVSKM
              250          260          270          280          290          300

orf7a      DGTGLSQFSHDLTEHNAAVRKYILKKX
              310          320          330

```

45 The complete length ORF7a nucleotide sequence (SEQ ID NO: 33) is:

```

1  ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
51 AGCCGTTTTTC GCCGCGCTGC TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTCGTCCGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
5  201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCCGCGT TTTCGCATAT
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTCAGCGGC
10 451 AATCCTGAAG GGCAGTTTTT CCCCAGACAG TACGAAATCG ATGCGGGCGG
501 CAGCGATTTA CGGATTTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGGCATGA
15 651 AGCCGACCGC GACCATGTCT GTTCCGTCTT CGTCAACCGC CTGAAAATCG
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATCGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
20 951 GACCGAACAC AACGCCGCCG TTCGCAAATA TATTTTGAAA AAATAA

```

This is predicted to encode a protein having amino acid sequence (SEQ ID NO: 34):

```

1  MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
25 51 LAEDRIVFSR HVLTAAYVL GVHNLHTGT YRLPSEVSAW DILQKMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP
201 YEMLIMASLI EKETGHEADR DHVASVFNRL KIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYL
30 301 FVSKMDGTGL SQFSDLTEH NAAVRKYILK K*

```

A leader peptide is underlined.

ORF7a (SEQ ID NO: 34) and ORF7-1 (SEQ ID NO: 32) show 98.8% identity in 331 aa overlap:

```

35 orf7a.pep      10      20      30      40      50      60
      MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
      |||||
orf7-1      MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
      10      20      30      40      50      60

40 orf7a.pep      70      80      90      100     110     120
      HVLTAAYVLGVHNLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
      |||||
orf7-1      HVLTAAYVLGVHNLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
      70      80      90      100     110     120

45 orf7a.pep      130     140     150     160     170     180
      IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
      |||||
orf7-1      IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM
      130     140     150     160     170     180

50 orf7a.pep      190     200     210     220     230     240
      QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTD
      |||||
orf7-1      QRRLEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFNRLKIGMRLQTD

```

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf7a.pep	PSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLY					
	orf7-1	PSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLY					
		250	260	270	280	290	300
		310	320	330			
10	orf7a.pep	FVSKMDGTGLSQFSHDLTEHNAAVRKYILKKX					
	orf7-1	FVSKMDGTGLSQFSHDLTEHNAAVRKYILKKX					
		310	320	330			

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 (SEQ ID NO: 30) shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) (SEQ ID NO: 36) from *N. gonorrhoeae*:

15	orf7	MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
	orf7ng	MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ	60
	orf7	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMAXLVEKETG	120
20	orf7ng	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG	120
	orf7	HEAXXDHVASVFNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRDTPYNTYTRGGLP	180
	orf7ng	HEADRDHVASVFNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTGGGLP	180
25	orf7	PTPIALP	187
	orf7ng	PTPIALPGKAAMDAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK	236

An ORF7ng nucleotide sequence (SEQ ID NO: 35) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 36):

30	1	MRGGRPDSVT	VQIIEGSRFS	HMRKVIDATP	DIGHDTKGWS	NEKLMAEVAP
	51	DAFSGNPEGQ	FFPDSYEIDA	GGSDLQIYQT	AYKAMQRRLN	EAWAGRQDGL
	101	PYKNPYEMLI	MASLIEKETG	HEADRDHVAS	VFNRLKIGM	RLQTDPSVIY
	151	GMGAAYKGKI	RKADLRDTP	YNTYTGGGLP	PTPIALPGKA	AMDAAHPSG
35	201	EKYLYFVSKM	DGTGLSQFSH	DLTEHNAAVR	KYILKK*	

Further sequence analysis revealed a partial DNA sequence of ORF7ng (SEQ ID NO: 37):

	1	..taccgaatca	AGATTGCCAA	AAATCAGGGT	ATTCGTCGG	TCGGCAGGAA
	51	ACTTGCCgaA	GACCGCATCG	TGTTCAAGCAG	GCATGTTTGT	ACAGCGGCGG
	101	CCTACGTTTT	GGGTGTGCAC	AACAGGCTGC	ATACGGGGAC	gTACAGATTG
40	151	CCTTCGGAAG	TGTCTGCTTG	GGATATCTTG	CAGAAAATGC	GCGGCGGCAG
	201	GCCGGATTCC	GTTACCGTGC	AGATTATCGA	AGGTTGCGGT	TTTTCGCATA
	251	TGAGGAAAGT	CATCGACGCA	ACGCCCCACA	TCGGACACGA	CACCAAAGGC
	301	TGGAGCAATG	AAAAACTGAT	GGCGGAAGTT	GCGCCCGATG	CCTTCAGCGG

5  
10

```

351 CAATCCTGAA GGGCAGTTTT TTCCCGACAG CTACGAAATC GATGCGGGCG
401 GCAGCGATTT GCAGATTTAC CAAACCGCCT ACAAGGCGAT GCAACGCCGC
451 CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC
501 TTATGAAATG CTGATTATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG
551 AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAAATC
601 GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC
651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA
701 aCAcTAtac gggcgggggc ttgccgcaa cccggattgc gctgccggc
751 Aaggcgcaa tggatgccgc cgcccaccg tccggcgaAa aatacctgTa
801 tttcgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
851 TGACCGAACA CAACGCCGcC gTcCGCAAAT ATATTTTGAA AAAATAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 38; ORF7ng-1):

15  
20

```

1  .YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
51  PSEVSAWDIL QKMRGGRPDS VTVQIIEGSR FSHMRKVIDA TPDIGHDTKG
101 WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR
151 LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI
201 GMRLQTDPSV IYGMGAAYKG KIRKADLRRD TPYNTYTGGG LPPTRIALPG
251 KAAMDAAHP SGEKYLYFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK*

```

ORF7ng-1 (SEQ ID NO: 38) and ORF7-1 (SEQ ID NO: 32) show 98.0% identity in 298 aa overlap:

25  
30  
35  
40  
45  
50

```

              10      20      30      40      50      60
orf7-1.pep    KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSRHVL
orf7ng-1      |||||
              10      20      30

              70      80      90      100     110     120
orf7-1.pep    TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDA
orf7ng-1      |||||
              40      50      60      70      80      90

              130     140     150     160     170     180
orf7-1.pep    TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR
orf7ng-1      |||||
              100     110     120     130     140     150

              190     200     210     220     230     240
orf7-1.pep    LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFNRLKIGMRLQTDPSV
orf7ng-1      |||||
              160     170     180     190     200     210

              250     260     270     280     290     300
orf7-1.pep    IYGMGAAYKGKIRKADLRRDTPYNTYTGGGLPPTPIALPGKAALDAAHPSGEKYLYFVS
orf7ng-1      |||||
              220     230     240     250     260     270

              310     320     330
orf7-1.pep    KMDGTGLSQFSHDLTEHNAAVRKYILKKX
              |||||

```



orf7ng-1 KMDGTGLSQFSHDLTEHNAAVRKYILKKX  
280 290

In addition, ORF7ng-1 (SEQ ID NO: 38) shows significant homology with a hypothetical *E.coli* protein (SEQ ID NO: 1114):

5 sp|P28306|YCEG\_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION  
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG\_ECOLI SW: P28306 but  
has 97 additional C-terminal residues [Escherichia coli] Length = 340  
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
Identities = 20/87 (22%), Positives = 40/87 (45%)

10 Query: 10 GISSVGRKLAEDRIVFSRHVLTAAYVLGVHNLHTGTYRLPSEVSAWDILQKMRGGRPD 69  
G ++G +L D+I+ V + + GTYR +++ ++L+ + G+  
Sbjct: 49 GRLALGEQLYADKIINRPRVFQWLLRIEPLDSHFKAGTYRFTPQMTVREMLKLLESGKEA 108

15 Query: 70 SVTVQIIIEGSRFSHMRKVIDATPDIGH 96  
++++EG R S K + P I H  
Sbjct: 109 QFPLRLVEGMRLSDYKQLREAPYIKH 135

Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
Identities = 84/155 (54%), Positives = 111/155 (71%)

20 Query: 120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRNLNEAWAGRQDGLPYKNPYEMLIMASLIEK 179  
EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK  
Sbjct: 158 EGWFWDPTWMYTANTTDVALLKRAHKMKVKAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217

Query: 180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKKGKIRKADLRDTPYNTYTGG 239  
ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT  
Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

25 Query: 240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274  
GLPP IA PG ++ AAAHP+ YLYFV+ G  
Sbjct: 278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 30 Example 6

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 39):

35 1 CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT  
51 ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG  
101 GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC  
151 GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC  
201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGCAGGCG GGAACGGCTC  
251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCG  
301 GAACGCGCCT TGGAAATGGC CGTGTCGCTG AACGCGTTG AACAGGCGGA  
351 AATGATTTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC

401 AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG  
 451 CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG

This corresponds to the amino acid sequence (SEQ ID NO: 40; ORF9):

5           1   ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRKQ QRYSEEEIKN  
           51   ERARLAAVGE RVNQIFTLG GETALQKQA GTALATYMLM LERTKSPEVA  
          101   ERALEMAVSL NAFEQAEMII QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ  
          151   HLDGREEVLA QADEGQ

10 Further sequence analysis revealed the complete DNA sequence (SEQ ID NO: 41):

1   ATGTTACCTA ACCGTTTCAA AATGTTAACT GTGTTGACGG CAACCTTGAT  
 51   TGCCGGACAG GTATCTGCCG CCGGAGGCGG TGCCGGGGAT ATGAAACAGC  
 101   CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA  
 151   GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA  
 15   201   TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGCAGG  
 251   CGGGAACGGC TCTGGCAACC TATATGCTGA TGTGGAACG CACAAAATCC  
 301   CCCGAAGTCG CCGAACGCGC CTTGGAAATG GCCGTGTCGC TGAACCGGTT  
 351   TGAACAGGCG GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC  
 401   CGGGTAAGGC GCAAAAACGG GCGGGGTGGC TGCGGAACGT GCTGAGGGAA  
 20   451   AGAGGAAATC AGCATCTGGA CCGACTGGAA GAAGTGCTGG CTCAGGCGGA  
 501   CGAAGGACAG AACC GCAGGG TGTTTTTATT GTTGGCACAA GCCGCGTGC  
 551   AACAGGACGG GTTGGCGCAA AAAGCATCGA AAGCGGTTCTG CCGCGCGGCG  
 601   TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTTCA  
 651   CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG  
 25   701   CGAAGCTCGA TACGGAATA TTGCCCCCA CTTTAATGAC GTTGCGTCTG  
 751   ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA  
 801   CACCCAAAAC CTTTCGGCCG TCTGGCAGGA AATGGAAATT ATGAATCTGG  
 851   TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTTG  
 901   GAACGCAATC CGAATGCAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC  
 30   951   AAACCGAAAA GAAGGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT  
 1001   ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGGCG  
 1051   ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA  
 1101   AAAAGTATCC GCGCCGGAAT ACCTGTTTCA CAAAGGTGTG CTGGCGGCTG  
 1151   CGGCGGCTGT CGAGTTGGAC GCGGCGCAGG CGGCTTTGCG GCAGATCGGC  
 35   1201   AGGGTGCGGA AACTTCCCGA ACAGCAGGGG CGGTATTTTA CGGCAGACAA  
 1251   TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG  
 1301   AGGCTTTGAG GGGGTTGGAC AAGATTATCG AAAAACC GCC TGCCGCGAGT  
 1351   AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTCAG TTGTTTACGA  
 1401   TCGGCTTGGC AAGCGGAAAA AAATGATTTT AGATCTTGAA AGGGCGTTCA  
 40   1451   GGCTTGACCC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG  
 1501   CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC  
 1551   ATACCAATC AACC CGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG  
 1601   CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT  
 1651   TCGTTTGAAA ACGACCCCGA GCCCGAAGTT GCCGCCCAT TGGGCGAAGT  
 45   1701   GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTGACGTA TGGACGCAGG  
 1751   CGGCACACCT TACGGGAGAC AAGAAAATAT GGCGGGAAAC GCTCAAACGT  
 1801   CACGGCATCG CATTGCCCA ACCTTCCCGA AAACCTCGGA AATAA

This corresponds to the amino acid sequence (SEQ ID NO: 42; ORF9-1):

50           1   MLPNRFKMLT VLTATLIAGQ VSAAGGGAGD MKQPKEVGKV FRKQQRYS  
           51   EIKNERARLA AVGERVNIQIF TLLGGETALQ KGQAGTALAT YMLMLERTKS  
          101   PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGKAQKR AGWLRNVLRE  
          151   RGNQHLDGLE EVLAQADEGQ NRRVFLLLAQ AAVQDGLAQ KASKAVRRRA  
          201   LKYEHLPEAA VADVVFVQG REKEKAIGAL QRLAKLDTEI LPPTLMTLRL

251 TARKYPEILD GFFEQTDTQN LSAVWQEMEI MNLVSLHRLD DAYARLNVLL  
 301 ERNPADLYI QAAILAANRK EGASVIDGYA EKAYGRGTEE QRSRAALTA  
 351 MMYADRRDYA KVRQWLKKVS APEYLFDKGV LAAAAVELD GGRAALRQIG  
 401 RVRKLPEQQG RYFTADNLSK IQMLALSKLP DKREALRGLD KIEKPPAGS  
 451 NTELQAEALV QRSVVYDRLG KRKKMISDLE RAFRLAPDNA QIMNMLGYSL  
 501 LTDSKRLDEG FALLQTAYQI NPDDTAVNDS IGWAYYLKGD AESALPYLRY  
 551 SFENDPEPEV AAHLGEVLWA LGERDQAVDV WTQAAHLTGD KKIWRETLKR  
 601 HGIALPQPSR KPRK\*

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF9 (SEQ ID NO: 40) shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) (SEQ ID NO: 44) from strain A of *N. meningitidis*:

15	orf9.pep	10 20 30 40 50	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFVRKQQRYSSEEEIKNERARLA
	orf9a	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFVRKQQRYSSEEEIKNERARLA
20	orf9.pep	60 70 80 90 100 110	AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9a	60 70 80 90 100 110	AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
25	orf9.pep	120 130 140 150 160	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ
	orf9a	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRVFLLLAQ
30	orf9a	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRRAALRYEHLPEAAVADVVSQXREKEKAIGALQRLAKLDTEI

The complete length ORF9a nucleotide sequence (SEQ ID NO: 43) is:

1 ATGTTACCCG CCCGTTTCAC CATTTTATCT GTGCTCGCGG CAGCCCTGCT  
 51 TGCCGGGCAG GCGTATGCCG CCGGCGCGGC GGATGCGAAG CCGCCGAAGG  
 35 101 AAGTCGGAAA GGTTTTTCAGA AAGCAGCAGC GTTACAGCGA GGAAGAAATC  
 151 AAAAACGAAC GCGCACGGCT TGCGGCAGTG GCGGAGCGGG TTAATCAGAT  
 201 ATTTACGTTG CTGGGANGGG AAACCGCCTT GCAAAAGGGG CAGGCGGGAA  
 251 CGGCTCTGGC AACCTATATG CTGATGTTGG AACGCACAAA ATCCCCGAA  
 301 GTCGCCGAAC GCGCCTTGGA AATGGCCGTG TCNCTGAACG CGTTTGAACA  
 40 351 GCGGAAATG ATTTATCAGA AATGGCGGCA GATTGAGCCT ATACCGGGTA  
 401 AGGCGCAAAA ACGGGCGGGG TGGCTGCGGA ACGTGCTGAG GGAAAGAGGA  
 451 AATCAGCATC TAGACGGAAT GGAAGAANTG CTGGCTCAGG CCGACGAANG  
 501 ACAGAACCGC AGGGTGT TTT TATTGTTGGC ACAAGCCGCC GTGCAACAGG  
 551 ACGGGTTGGC GCAAAAAGCA TCGAAAGCGG TTCGCCGCGC GCGGTTGAGA  
 45 601 TATGAACATC TGCCCGAAGC GCGGTTGCC GATGTGGTGT TCAGCGTACA  
 651 GGNACGCGAA AAGGAAAAGG CAATCGGAGC TTTGCAGCGT TTGGCGAAGC  
 701 TCGATACGGA AATATTGCCC CCCACTTTAA TGACGTTGCG TCTGACTGCA  
 751 CGCAAATATC CCGAAATACT CGACGGCTTT TTCGAGCAGA CAGACACCCA

5  
10  
15  
20

```

801 AAACCTTTTCG GCCGTCTGGC AGGAAATGGA AATTATGAAT CTGGTTTCCC
851 TGCACAGGCT GGATGATGCC TATGCGCGTT TGAACGTGCT GTTGGAACGC
901 AATCCGAATG CAGACCTGTA TATTCAGGCA GCGATATTGG CGGCAAACCG
951 AAAAGAANGT GCTTCCGTTA TCGACGGCTA CGCCGAAAAG GCATACGGCA
1001 GGGGACGGG GGAACAGCGG GGCAGGGCGG CAATGACGGC GGCGATGATA
1051 TATGCCGACC GAAGGGATTA CACCAAAGTC AGGCAGTGGT TGAAAAAAGT
1101 GTCCGCGCCG GAATACCTGT TCGACAAAGG TGTGCTGGCG GCTGCGGCGG
1151 CTGTCTGAGTT GGACNGCGGC AGGGCGGCTT TCGGCGAGAT CGGCAGGGTG
1201 CGGAAACTTC CCGAACAGCA GGGGCGGTAT TTACGGCAG ACAATTTGTC
1251 CAAAATACAG ATGTTTCGCC TGTGGAAGCT GCGGACAAA CGGGAGGCTT
1301 TGAGGGGGTT GGACAAGATT ATCGAAAAAC CGCCTGCCGG CAGTAATACA
1351 GAGTTACAGG CAGAGGCATT GGTACAGCGG TCAGTTGTTT ACGATCGGCT
1401 TGGCAAGCGG AAAAAAATGA TTTAGATCTT TGAAAGGGCG TTCAGGCTTG
1451 CACCCGATAA CGCTCAGATT ATGAATAATC TGGGCTACAG CCTGCTTTCC
1501 GATTCCAAAC GTTTGGACGA AGGCTTCGCC CTGCTTCAGA CGGCATACCA
1551 AATCAACCCG GACGATACCG CTGTCAACGA CAGCATAGGC TGGGCGTATT
1601 ACCTGAAANG CGACGCGGAA AGCGCGCTGC CGTATCTGCG GTATTTCGTTT
1651 GAAAACGACC CCGAGCCCGA AGTTGCCGCC CATTTGGGCG AAGTGTGTG
1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
1751 ACCTTACGGG AGACAAGAAA ATATGGCGGG AAACGCTCAA ACGTCACGGC
1801 ATCGCATTGC CCCAACCTTC CCGAAAACCT CGGAAATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 44):

25  
30  
35

```

1  MLPARFTILS VLAAALLAGQ AYAAGAADAK PPKEVGKVFR KQQRYSSEEEI
51  KNERARLAAV GERVNQIFTL LGXETALQKG QAGTALATYM LMLERTKSPE
101 VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGKAQKRAG WLRNVLRERG
151 NQHLDGLEEX LAQADEXQNR RVFLLLAQAA VQDGLAQKA SKAVRRAALR
201 YEHLPEAAVA DVVFSVQXRE KEKAIGALQR LAKLDTEILP PTLMTLRLTA
251 RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLHRLDDA YARLNVLLER
301 NPNADLYIQA AILANRKEKX ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
351 YADRRDYTKV RQWLKKVSAP EYLPDKGVLA AAAAVELDXG RAALRQIGRV
401 RKLPEQQGRY FTADNLSKIQ MFALSKLPDK REALRGLDKI IEKPPAGSNT
451 ELQAEALVQR SVVYDRLGKR KKMISDLERA FRLAPDNAQI MNNLGYSLLS
501 DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLKXDAE SALPYLRYSF
551 ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLTGDKK IWRETLKRHG
601 IALPQPSRKP RK*

```

ORF9a (SEQ ID NO: 44) and ORF9-1 (SEQ ID NO: 42) show 95.3% identity in 614 aa overlap:

40  
45  
50

```

          10      20      30      40      50
orf9a.pep  MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRKQQRYSSEEEIKNERARLA
          |||  ||  :||:|:|:|:|:|:|  |||  |:|  |  |||||
orf9-1     MLPNRfKMLTTLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQQRYSSEEEIKNERARLA
          10      20      30      40      50      60

          60      70      80      90      100     110
orf9a.pep  AVGERVNQIFTL LGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
          |||||
orf9-1     AVGERVNQIFTL LGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
          70      80      90      100     110     120

          120     130     140     150     160     170
orf9a.pep  EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRVFLLLAQ
          |||||
orf9-1     EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRVFLLLAQ
          130     140     150     160     170     180

```

5	orf9a.pep	180	190	200	210	220	230
	orf9-1	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQXREKEKAIGALQRLAKLDTEI	190	200	210	220	230
10	orf9a.pep	240	250	260	270	280	290
	orf9-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL	250	260	270	280	290
15	orf9a.pep	300	310	320	330	340	350
	orf9-1	ERNPNADLYIQAAILAANRKEKASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYT	310	320	330	340	350
20	orf9a.pep	360	370	380	390	400	410
	orf9-1	KVRQWLKKVSAPEYLFDKGVLA AAAA VELDXGRAALRQIGRVRKLPEQQGRYFTADNLSK	370	380	390	400	410
25	orf9a.pep	420	430	440	450	460	470
	orf9-1	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTTELQAEALVQRSVVYDRLGKRKKMISDLE	430	440	450	460	470
30	orf9a.pep	480	490	500	510	520	530
	orf9-1	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD	490	500	510	520	530
35	orf9a.pep	540	550	560	570	580	590
	orf9-1	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR	550	560	570	580	590
40	orf9a.pep	600	610				
	orf9-1	HGIALPQPSRKPRKX	610				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 (SEQ ID NO: 40) shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) (SEQ ID NO: 46) from *N. gonorrhoeae*:

45	Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFVRKQQRYS EEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLKRHRYS EEEIKNERAR	58

orf9	LAAVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
orf9	QAEMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
orf9ng	QAEMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL	178

The ORF9ng nucleotide sequence (SEQ ID NO: 45) was predicted to encode a protein having including acid sequence (SEQ ID NO: 46):

1	MIMLPARFTI	LSVLAAALLA	QAYAAAGAAD	VELPKEVGKV	LRKHRRYSEE
51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIPGEAQKP	AGWLRNVLKE
151	GGNPHLDRLE	EVPAQSDYVH	QPMIFLLLVQ	AAVQHGGVAQ	KPSKAVRPAA
201	YNYEVLPEA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
251	RPTARISP	LLQRFRTPEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

Further sequence analysis revealed the complete length ORF9ng DNA sequence (SEQ ID NO: 47):

1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCTCTGCAG	CAGCCCTGCT
51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
101	AAGTCGGAAA	GGTTTTAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
25	251	CGGCTCTGGC	AACCTATATG	CTGATGTGG	AACGCACAAA
301	GTCCGCCAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
351	GGCGGAAATG	ATTTATCAGA	AATGgcggca	gatcgagcct	ataCcggggtg
401	aggcgcaaaa	accgGcgggG	tggctgcgga	acgtattgaa	ggaagggGGa
451	aaTCAGCATC	TGGAcgggtt	gaaagaggTG	CtggcgcaAT	cggacgatGT
501	GCAAAAacgc	aggaTATTTT	TGCTGTGTTT	GCAAGCCGCC	GTGcagcagg
551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTCGCcgtgc	GGcgttgaAG
601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGcAGCGT	TTGGCGAAGC
701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCA
801	AAACCTTTCG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACAC
901	AAACCGAATG	CAAACCTGTA	TATTCAGGCG	GCGATATTGG	CGGCAAACCG
951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	caATgacggc	GGCGATGATA
1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGGCGG
1151	CTGCCGAATT	GGACGGAGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTGTCT
1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAAGCCC
1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCGGCGGC	GGGAAGCACG
1351	GAACCTTTGG	CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
1401	cggCAAACGG	GGAAAAATGA	TTGCCGACCT	tgaAACcgcg	CTCAAACCTTA
1451	CGCCCGATAA	TGCACAAATT	ATGAATAATC	TGGGTACAG	CCTGCTTTCC
1501	GATTCCAAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
1551	AATCAACCCG	GACGATACCG	CCGTTAACGA	CAGCATAGGC	TGGGCGTATT



		300	310	320	330	340	350
			370	380	390	400	410
	orf9-1.pep	KVRQWLKKVSAPEYLFDPKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK					
5	orf9ng-1	KVRQWLKKVSAPEYLFDPKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK					
		360	370	380	390	400	410
			430	440	450	460	470
	orf9-1.pep	IQMLALSKLPDKREALRGLDKII EKPPAGSNTTELQAEALVQRSVVYDRLGKRKKMISDLE					
10	orf9ng-1	IQMLALSKLPDKREALIGLNNII AKLSAAGSTEPLAEALQRSIIYEQFGKRGMIAADLE					
		420	430	440	450	460	470
			490	500	510	520	530
	orf9-1.pep	RAFRLAPDNAQIMNNGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLGKD					
15	orf9ng-1	TALKLTPDNAQIMNNGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLGKD					
		480	490	500	510	520	530
			550	560	570	580	590
	orf9-1.pep	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRET LKR					
20	orf9ng-1	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRET LKR					
		540	550	560	570	580	590
			610				
	orf9-1.pep	HGIALPQPSRKPRKX					
25	orf9ng-1	YGIALPEPSRKPRKX					
		600	610				

In addition, ORF9ng (SEQ ID NO: 48) shows significant homology with a hypothetical protein (SEQ ID NO: 1115) from *P.aeruginosa*:

30	sp P42810 YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION (ORF3)
	)gi 1072999 pir  S49376 hypothetical protein 3 - Pseudomonas aeruginosa )gi 557259 (X82071) orf3 [Pseudomonas aeruginosa] Length = 576
35	Score = 128 bits (318), Expect = 1e-28
	Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)
	Query: 67 VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIIYQKWR 126
	+++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
	Sbjct: 53 LYSLLVAELAGQRNRFDIALSNYVVQAKTRDPGVSERAFRIA EYL GADQEALDTSLLWA 112
40	Query: 127 QIEPIPGEAQKPAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRRI 172
	+ P +AQ+ A ++ VL G+ H D L A++D + +
	Sbjct: 113 RSAPDNLDAQRAAIIQLARAGRYEESMVYMEKVLNGQGDTHFDLALSAAETDPDTRAGL 172
	Query: 173 FXXXXXXXXXXXXXXXXKASKAVRRRAALKY EHLPEAAVADAVFGVQGREKEKAIEALQRLA 232
	++ KY + + A+ Q ++A+ L+ +
	Sbjct: 173 L-----QSF D HLLKKYPNNGQLLFGKALLLQQDGRPD EALTLEDNS 214
45	Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKP 287
	E+ P L + L + K P + G E D + + + + LV +
	Sbjct: 215 ASRHEVAPLLLRSLRLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270



Query: 288 DDAYARLNVLEHNPN-----ANLYIQAAI----- 312  
 DDA A L++ P+ A +Y++ +  
 Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFSALVCLQAQWDEARIYLEELVERDSHVDAAHFNLG 330

5 Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371  
 LA +K+ A +D YA+ G G + T ++ A R D A R + P+  
 Sbjct: 331 RLAEQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

Query: 372 YLFDKXXXXXXXXXXXXXXXXXXXXXQIGRVRKLPEQQGRYFTADNLSKIQMLALSCLPDKR 431  
 Y A L I+ ALS +  
 Sbjct: 389 Y-----AIQYLIIEAEALSNNDDQOE 408

10 Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491  
 +A + + + E L L RS++ E+ +M DL + PDNA +  
 Sbjct: 409 KAWQAIQEGLKQYP----EDL-NLLYTRSMLEKRNDLAQMEKDLRFVIAREPDNAMAL 462

15 Query: 492 NNLGYSLLSDSKRLDEGFALLQYATQINPDDTAVNDSIGWAYYLKGDAESALPYLRYSFE 551  
 N LGY+L + R E L+ A+++NPDD A+ DS+GW Y +G A YLR + +  
 Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILDSMGWINYRQGLADAERYLRQALQ 522

Query: 552 NDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR 598  
 P+ EVAHLGEVLWA G + A +W + + D + R T+KR  
 Sbjct: 523 RYPDHEVA AHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

20 gi|2983399 (AE000710) hypothetical protein (SEQ ID NO: 1116) [Aquifex aeolicus]  
 Length = 545  
 Score = 81.5 bits (198), Expect = 1e-14  
 Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

25 Query: 408 GRYFTADNL-SKIQMLALSCLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQ----- 459  
 G Y A L K ++LA PDK+E L + +K + + L +  
 Sbjct: 335 GNYEDAKRLIEKAKVLA---PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513  
 +I+Y+ G L A++L P+N N LGYSLL +R++E L++  
 Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYNLGYSLLLWYGKERVEEAEELIK 450

30 Query: 514 TAYQINPDDTAVNDSIGWAYYLKGDAESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572  
 A + +P++ A DS+GW YYLKG D E A+ YL + E +P V H+G+VL +G +  
 Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKG DYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510

Query: 573 DQAVDVWTQAAHLRGDKK 590  
 ++A + + +A L + K  
 Sbjct: 511 EARNYYERALKLLEEGK 528

35 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 7

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 49):

40 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA  
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

```

101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
151 TGGGCGATTA TCGTTTAAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGaCTGGGCG
351 GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
401 GCATTGTTTCG CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT
451 TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG
501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCGCGCCG GAcCGACCCG
551 ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTgTT
601 CTTCTTCTTC CCTGCCGgks TGGTATTGTA CTGGGTAGTC AACAACCTCC
651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
701 GCCCAAGGCG AAGTCGTTTC CTA

```

15 This corresponds to the amino acid sequence (SEQ ID NO: 50; ORF11):

20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160

```

1  .NLYAGPQTTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG
51  WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ
101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGI
151 TDLRADPYI ILPIIMAATM FAQTYLNPPP TDPMQAKMMK IMPLVFSXXF
201 FFFPAGXVLY WVNNLLTIA QQWHINRSIE KRAQGEVVS *

```

Further sequence analysis revealed the complete DNA sequence (SEQ ID NO: 51):

25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160

```

1  ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT
51  GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCGGCCGA AGCCGCGCTC
151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
201 TGATGAAAAA AGCGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGCGA CGGCAAAAGAA
301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG
401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
451 ATCGACAAAG TTTTACTTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
501 CTTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
551 ACCGATCGT CCGGACCCAC AGCGAACCCG AGGGTCAAGG TTACTTTACC
601 CACTCTTACG TCGGCTCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
801 CGCCGCAAGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
851 ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA
901 GCCGAAGCCT CATCAACCT CTACGCGGCG CCGCAGACCA CATCCGTCAT
951 CGCAAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAAGTAC
1001 ACTGGTTCGC CTCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCTCTTA CCGCTCTATG GCGAAAATGC
1151 GTGCCCGCGC ACCCAAAGTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACACAGGC GATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCCGCTG GGCGGCTGCC TGCTATGCT GTTGCAAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTGCG CCAAACCTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAATATCAT GCCGTTGGTT
1501 TTCTCCGTCA TGTCTTCTT CTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA
1601 TCGAAAAACA ACGCGCCCAA GGCGAAGTCG TTCCTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 52; ORF11-1):

```

      1  MDFKRLTAFF AIALVIMIGW EKMFFTPKPV PAPQQAQQQ AVTASAEAL
    51  APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFILFGDGKE
   101  YTYVAQSELL DAQGNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
   151  IDKVYFTFTG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
   201  HSYVGPVVYT PEGNFQKVSF SDLDDDAKSG KSEAHEYIRKT PTGWLGMIEH
   251  HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
   301  AEASINLYAG PQTTSVIANI ADNLQLAKDY GKVHWFASPL FWLLNQLHNI
   351  IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEKYGDD
  10  401  RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
   451  WLGWITDLR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
  501  FSVMFFFFPA GLVLYWVVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*

```

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a 60kDa inner-membrane protein (accession P25754) (SEQ ID NO: 1117) of *Pseudomonas putida*

ORF11 (SEQ ID NO: 50) and the 60kDa protein (SEQ ID NO: 1117) show 58% aa identity in 229 aa overlap (BLASTp).

```

 20  ORF11  2  LYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIIVLTIIVK 61
      LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
      60K  324 LYAGPKIQSKLKELSPGLELTVDYDGLWFIAQPIFWLLQHIHSLGNGWWSIIIVLTMLIK 383

      ORF11  62  AVLYPLTNASYRSMAMRAAAPKLQAIKEKYGDDRRXXXXXXXXXXLYTDEKINPLGGCLPM 121
      + +PL+ ASYRSMA+MRA APKL A+KE++GDDR LY EKinPLGGCLP+
      60K  384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443

 25  ORF11  122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLRDPYYILPII MAATMFAQTYLNPPPT 181
      L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P
      60K  444 LVQMPVFLALYWVLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQRLNPTPP 503

      ORF11  182 DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYWVVNNLLTIAQQWHINRSIE 230
      DPMQAK+MK+MP++ PAG VLYWVVNN L+I+QQW+I R IE
 30  60K  504 DPMQAKVMKMMPPIIFTFFFLWFPAGLVLYWVVNNCLSSISQWYITRRIE 552

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 (SEQ ID NO: 50) shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) (SEQ ID NO: 54) from strain A of *N. meningitidis*:

```

 35  orf11.pep
      NLYAGPQTTSVIANIADNLQLAKDYGKVHW
      |||||
  orf11a  IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQLXKDYGKVHW
      280      290      300      310      320      330

```

-110-

		40	50	60	70	80	90
orf11.pep		FASPLFWLLNQLHNIIGNWGWAIIVLTIIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKE					
orf11a		FASPLFWLLNQLHNIIGNWGWAIIVLTIIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKE					
5		340	350	360	370	380	390
		100	110	120	130	140	150
orf11.pep		KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI					
orf11a		KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI					
10		400	410	420	430	440	450
		160	170	180	190	200	210
orf11.pep		TDLSRADPYYILPIIIMAATMFAQTYLNPPTDPMQAKMMKIMPLVFSXXXFFFPAGXVLY					
orf11a		TDLSRADPYYILPIIIMAATMFAQTYLNPPTDPMQAKMMKIMPLVXSSXXFFXFPAGLVLY					
15		460	470	480	490	500	510
		220	230	240			
orf11.pep		WVNNLLTIAQQWHINRSIEKQRAQGEVVSX					
orf11a		WVINLLTIAQQWHINRSIEKQRAQGEVVSX					
20		520	530	540			

The complete length ORF11a nucleotide sequence (SEQ ID NO: 53) is:

	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTAT
	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCCgcgcccc
25	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCGCGCGA	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTCAT
	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CNAAAATAAA	CCGTTCATCC	TGTTTGGCGA	CGGCAANAA
30	301	TACACCTACN	TCGCCCANTC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	AGCTTGGAAG
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCAC	CTGAAACACG	CGGTCTGAAA
	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	AGCTATCTGG	TCAACGTCCG
	501	CTTCGACATC	GCCAAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAAACCG	AGGGTCAAGG	CTACTTTACC
35	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAATA
	651	AGTCAGCTTC	TCCGACTTGG	ACGACGATGC	CAANTCCGGN	AAATCCGAGG
	701	CCGAATACAT	CCGCAAAACC	CNGACCGGCT	GGCTCGGCAT	GATTGAACAC
	751	CACTTCATGT	CCACCTGGAT	CCTCCAACCC	AAAGGCGGAC	AAAGCGTTTG
40	801	CGCCGCTGGC	GACTGCNGTA	TNGACATCAA	ACGCCGCAAC	GACAAGCTGT
	851	ACAGCACCAG	CGTCAGCGTG	CCTTTAGCCG	CTATCCAAAA	CGGTGCGAAA
	901	TCCNAAAGCCT	CCATCAACCT	CTACGCCGGC	CCACAGACCA	CATCNGTTAT
	951	CGCAACATC	GCCGACAACC	TGCAACTGGN	CAAAGACTAC	GGCAAAAGTAC
	1001	ACTGGTTCGC	CTCCCCCTC	TTTGGCTTTT	TGAACCAACT	GCACAACATC
45	1051	ATCGGCAACT	GGGGCTGGGC	GATTATCGTT	TTAACCATCA	TCGTCAAAGC
	1101	CGTACTGTAT	CCATTGACCA	ACGCCTCTTA	CCGTTCGATG	GCGAAAATGC
	1151	GTGCCGCCGC	GCCCAAACTG	CAAGCCATCA	AAGAGAAATA	CGCGGACGAC
	1201	CGTATGGCGC	AGCAACAAGC	CATGATGCAG	CTTTACACAG	ACGAGAAAAAT
	1251	CAACCCGCTG	GGCGGCTGCC	TGCCTATGCT	GTTGCAAATC	CCCGTCTTCA
50	1301	TCGGATTGTA	TTGGGCATTG	TTCCGCTCCG	TAGAATTGCG	CCAGGCACCT
	1351	TGGCTGGGTT	GGATTACCGA	CCTCAGCCGC	GCCGACCNT	ACTACATCCT
	1401	GCCCATCATT	ATGGCGGCAA	CGATGTTGCG	CCAAACCTAT	CTGAACCCGC
	1451	CGCCGACCGA	CCCGATGCAG	GCGAAAATGA	TGAAAATCAT	GCCTTTGGTT
	1501	NTNTCNNNNA	NGTTCCTTCNN	CTTCCTTGCC	GGTCTGGTAT	TGTACTGGGT
	1551	GATCAACAAC	CTCCTGACCA	TCGCCCAGCA	ATGGCACATC	AACCGCAGCA
55	1601	TCGAAAAACA	ACGCGCCCAA	GCGGAAGTCG	TTTCCTAA	

This encodes a protein having amino acid sequence (SEQ ID NO: 54):

```

1  XDFKRLTXFF AIALVIMIGX XXMFPTPKPV PAPQQTAAQQ AVXASAEAL
51 APXXPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDxNK PFILFGDGKX
101 YTYXAXSELL DAQNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVYFTFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVVYT PEGNFQKVSF SDLDDAXSG KSEAHEYIRKT XTWLGMIEH
251 HFMSTWILQP KGGQSVCAAG DCXXDIKRRN DKLYSTSVSV PLAAIQNGAK
301 SXASINLYAG PQTTSVIANI ADNLQLXKDY GKVHWFASPL FWLLNQLHNI
10 351 IGWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAPKL QAIKEYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLSR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501 XSXXFFXFPF GLVLYWVINN LLTIAQQWHI NRSIEKQRAQ GEVVS*

```

15 ORF11a (SEQ ID NO: 54) and ORF11-1 (SEQ ID NO: 52) show 95.2% identity in 544 aa overlap:

		10	20	30	40	50	60
orf11a.pep		XDFKRLTXFFAIALVIMIGXXXMFPTPKVP	PAPQQTAAQQAVXASAEALAPXXPITVTT				
orf11-1		MDFKRLTAFFAIALVIMIGWEKMFPTPKVP	PAPQQAQQAVTASAEALAPATPITVTT				
		10	20	30	40	50	60
		70	80	90	100	110	120
orf11a.pep		DTVQAVIDEKSGDLRRLTLLKYKATGDxNKPFILFGDGKXYTYXAXSELLDAQNNILKG					
orf11-1		DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQNNILKG					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf11a.pep		IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYFTFTKGSYLVNVRFDIANGSGQTANL					
orf11-1		IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYFTFTKGSYLVNVRFDIANGSGQTANL					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf11a.pep		SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAXSGKSEAHEYIRKT					
orf11-1		SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAHEYIRKT					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf11a.pep		XTWLGMIIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK					
orf11-1		PTWLGMIIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK					
		250	260	270	280	290	300
		310	320	330	340	350	360
orf11a.pep		SXASINLYAGPQTTSVIANIADNLQLXKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV					
orf11-1		AEASINLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV					
		310	320	330	340	350	360
		370	380	390	400	410	420
orf11a.pep		LTIIVKAVLYPLTNASYRSMAMKRAAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINPL					
orf11-1		LTIIVKAVLYPLTNASYRSMAMKRAAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINPL					
		370	380	390	400	410	420

		430	440	450	460	470	480
	orf11a.pep	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPII MAATMFAQTY					
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPII MAATMFAQTY					
5		430	440	450	460	470	480
	orf11a.pep	LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ					
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ					
10		490	500	510	520	530	540
	orf11a.pep	LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ					
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ					
		490	500	510	520	530	540
	orf11a.pep	GEVVSX					
	orf11-1	GEVVSX					

# 15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 (SEQ ID NO: 50) shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) (SEQ ID NO: 56) from *N. gonorrhoeae*:

	Orf11	NLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIVVLT	60
20	orf11	IIVKAVLYPLTNASYRSMKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMKMRAAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
25	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPII MAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPII MAATMFAQTYLN	180
	orf11	PPPTDPMQAKMMKIMPLVFSXXFFFPAGXVLYWVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGE	240
30	orf11	VVS 240	
	orf11ng	VVS 243	

An ORF11ng nucleotide sequence (SEQ ID NO: 55) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 56):

	1	MAVNLYAGPQ	TTSVIANIAD	NLQAKDYGK	VHWFASPLFW	LLNQLHNIIG
	51	NWGWAIVVLT	IIVKAVLYPL	TNASYRSMK	MRAAAPQLT	IKEKYGDDRM
	101	AQQQAMMQLF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELRQAPWL
	151	GWITDLSRAD	PYYILPII MA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
40	201	VMFFFFPAGL	VLYWVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence (SEQ ID NO: 57) to be:

```

      1  ATGGATTTTA AAAGACTCAC GGCGTTTTTC GCCATCGCGC TGGTGATTAT
      51  GATCGGCTGG GAAAAAATGT TCCCCACCCC GAAACCCGTC CCCGCGCCCC
5    101  AACAGGCGGC AAAAAACAG GCAGCAACCG CTTCGCGCGA AGCCGCGCTC
      151  GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTTAT
      201  TGATGAAAAA AGTGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
      251  CAACCGGCGA CGAAACAAA CCGTTCGTCC TGTTTGGCGA CGGCAAAAGAA
      301  TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
10   351  TCTGAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC ACCCTCAACG
      401  GCGACACAGT CGAAGTCCGC CTGAGCGCGC CCGAAACCAA CGGACTGAAA
      451  ATCGACAAAG TCTATACCTT TACCAAAGAC AGCTATCTGG TCAACGTCCG
      501  CTTGACATC GCCAACGGCA GCGGTCAAAAC CGCCAACCTG AGCGCGGACT
      551  ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG CTACTTTACC
15   601  CACTTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
      651  AGTCAGCTTC TCCgacTTgg acgACGATGC gaaaTccggc aaATccgagg
      701  ccgaatacaT CCGCAAAACC ccgaccggtt ggctcggcat gattgaacac
      751  cacttcatgt ccacctggat cctccAAcct aaaggcggcc aaaacgtttg
      801  cgcccagggg gactgccgta tcgacattaa aCgccgcaac gacaagctgt
20   851  acagcgcaag cgtcagcgtg cctttaaccg ctatcccaac ccggggggcca
      901  aaaccgaaaa tggcggTCAA CCTGTATGCC GGTCCGCAAA CCACATCCGT
      951  TATCGCAAAC ATCGCcgacA ACCTGCAACT GGCAAAAGAC TACGGTAAAG
1001  TACACTGGTT CGCATCGCCG CTCTTCTGGC TCCTGAACCA ACTGCACAAC
25   1051  ATTATCGGCA ACTGGGGCTG GGCAATCGTC GTTTTGACCA TCATCGTCAA
      1101  AGCCGTACTG TATCCATTGA CCAACGcctc ctACCGTTTCG ATGGCGAAAA
      1151  TGCGTGccgc cgcacCcaaa CTGCAGACCA TCAAAGAAAA ATAcgGCGAC
      1201  GACCGTATGG CGCAACAGCA AGCGATGATG CAGCTTTACA AAgacgAGAA
      1251  AATCAACCCG CTGGGCGGCT GTctgcctat gctgttgCAA ATCCCGTCT
30   1301  TCATCGGCTT GTACTGGGCA TTGTTCGCCT CCGTAGAATT GCGCCAGGCA
      1351  CCTTGGCTGG GCTGGATTAC CGACCTCAGC CGCGCCGACC CCTACTACAT
      1401  CCTGCCCATC ATTATGGCGG CAACGATGTT CGCCCAAACC TATCTGAACC
      1451  CGCCGCCGAC CGACCCGATG CAGGCGAAAA TGATGAAAT CATGCCGTTG
      1501  GTTTTCTCCG TCATGTTCTT CTTCTTCCCT GCCGGTTTGG TTCTCTACTG
35   1551  GGTGGTCAAC AACCTCTGA CCATCGCCCA GCAGTGGCAC ATCAACCGCA
      1601  GCATCGAAAA ACAACGCGCC CAAGGCGAAG TCGTTTCCTA A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 58; ORF11ng-1):

```

      1  MDFKRLTAFF AIALVMIGW EKMFPPTPKPV PAPQQAQKQ AATASAEAL
      51  APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFVLFGDGKE
40   101  YTYVAQSELL DAQGNILKG IGFSAPKKQY TLNGDTVEVR LSAPETNGLK
      151  IDKVYTFTKD SYLVNVRFDI ANSGQTANL SADYRIVRDH SEPEGQGYFT
      201  HSYVGPVVTY PEGNFQKVSF SLDLDDAKSG KSEAEYIRKT PTGWLGMIEH
      251  HFMSTWILQP KGGQNVCAQG DCRIDIKRRN DKLYSASVSV PLTAIPTRGP
      301  KPKMAVNLYA GPQTTSVIAN IADNLQLAKD YGKVHWFASP LFWLLNQLHN
45   351  IIGNWGWAIIV VLTIIKAVL YPLTNASYRS MAKMRAAPK LQTIKEKYGD
      401  DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA
      451  PWLGWITDLS RADPYILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
      501  VFSVMFFFFP AGLVLYWVVN NLLTIAQQWH INRSIEKQRA QGEVVS*

```

50 ORF11ng-1 (SEQ ID NO: 58) and ORF11-1 (SEQ ID NO: 52) shown 95.1% identity in 546 aa overlap:

5	orf11ng-1.pep	10	20	30	40	50	60
	orf11-1	10	20	30	40	50	60
10	orf11ng-1.pep	70	80	90	100	110	120
	orf11-1	70	80	90	100	110	120
15	orf11ng-1.pep	130	140	150	160	170	180
	orf11-1	130	140	150	160	170	180
20	orf11ng-1.pep	190	200	210	220	230	240
	orf11-1	190	200	210	220	230	240
25	orf11ng-1.pep	250	260	270	280	290	300
	orf11-1	250	260	270	280	290	300
30	orf11ng-1.pep	310	320	330	340	350	360
	orf11-1	300	310	320	330	340	350
35	orf11ng-1.pep	370	380	390	400	410	420
	orf11-1	360	370	380	390	400	410
40	orf11ng-1.pep	430	440	450	460	470	480
	orf11-1	420	430	440	450	460	470
45	orf11ng-1.pep	490	500	510	520	530	540
	orf11-1	480	490	500	510	520	530
50	orf11ng-1.pep	QGEVVSX					
	orf11-1	540					



In addition, ORF11ng-1 (SEQ ID NO: 58) shows significant homology with an inner-membrane protein from the database (accession number p25754) (SEQ ID NO: 1117):

ID 60IM\_PSEPU STANDARD; PRT; 560 AA.  
 AC P25754;  
 DT 01-MAY-1992 (REL. 22, CREATED)  
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE 60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES Init1: 1074 Initn: 1293 Opt: 1103  
 Smith-Waterman score: 1406; 41.5% identity in 574 aa overlap

```

      10      20      30      40
orf11ng-1.pep MDFKR---LTAFFAIALVIMIGW----EKMFFT-----PKPVPAQQAAQKQ
      ||:||  ::|: ::|: ||:|  :  :||  | |||  ::|: :
p25754      MDIKRTILIAALAVVSYVMVLKWNDDYGQAALPTQNTAASTVAPGLPDGVAGNNGASAD
      10      20      30      40      50      60

      50      60      70      80      90
orf11ng-1.pep AATASAEALAPATPIT-----VTTDTVQAVIDEKSGDLRRLTLLKYKATGDE-NKPF
      :|:|:|:|:|  :|:|  ||:|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|
p25754      VPSANAESSPAELAPVALSKDLIRVKTDVLELAIDPVGGDIVQLNLPKYPRRQDHPNIPF
      70      80      90      100     110     120

      100     110     120     130     140
orf11ng-1.pep VLFGDGKEYTYVAQSELLDAQGNNILKGIG---FSAPKKQYTL-NGD---TVEVRLSAPE
      || :|  :|:|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|:|
p25754      QLFDNNGGERVYLAQSGLTGTDGPDA-RASGRPLYAAEQKSYQLADGQEQLVVDLKFPS---
      130     140     150     160     170

      150     160     170     180     190     200
orf11ng-1.pep TNGLKIDKVYTFKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY
      ||:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|
p25754      DNGVNYIKRFSFKRGEYDLNVSYLIDNQSGQAWNGNMFAQLKRDASGDPSSSTATGTATY
      180     190     200     210     220     230

      210     220     230     240     250     260
orf11ng-1.pep VGPVVYTPEGNFQKVSFSDLDDAKSGKSEAERYKTPPTGWLGMIEHHFMSTWILQPKGG
      :|  ::|  :|:|:|:|:|  ||:|  :|  :|:|  :|:|  :|:|:|:|  :|
p25754      LGAALWTASEPYKKVSMKDID---KGSLKE-----NVSGGWAWLQHYFVTAWI-PAKSD
      240     250     260     270     280

      270     280     290     300     310     320
orf11ng-1.pep QNVCAQGDCRIDIKRRNDKLYSASVSVPLTAIPTRGPKPKMAVNLYAGPQTTSVIANIAD
      :||  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|  :|:|
p25754      NNV-----VQTRKDSQGYNIIGYTGPIVSVPA-GGKVETSALLYAGPKIQSKLKELSP
      290     300     310     320     330

      330     340     350     360     370     380
orf11ng-1.pep NLQLAKDYGKVHWF-ASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMA
      :|:|:|  |||  :||  :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
p25754      GLELTVDYGFL-WFIAQPIFWLLQHIHSLLGNGWWSIIVLTMLIKGLFFPLSAASYRSMA
      340     350     360     370     380     390

      390     400     410     420     430     440
orf11ng-1.pep KMRAAAPKLQTIKEKYGDDRMQAQQAMMQLYKDEKINPLGGCLPMLLIQIPVFIGLYWALF

```

```

p25754      :|||:|||| :|||:|||| :|||:|||| |||||:||||:|:|:||||:|:|:|:
              RMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYWVLL
              400      410      420      430      440      450

              450      460      470      480      490      500
5  orfl1ng-1.pep ASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVF
              |||:|||||: ||||| |||:|||||:|||| | ||| | |||||:|:|:|:|
p25754      ESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQRLNPTPPDPMQAKVMKMMPIIF
              460      470      480      490      500      510

              510      520      530      540
10 orfl1ng-1.pep SVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGEVVSX
              :|:|:| ||||| ||||| |:|:| |||:| ||
p25754      TFFFLWFPAGLVLYWVVNNCLSSISQQWYITRRIEAATKKAAA
              520      530      540      550      560

```

- 15 Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonoccal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 8

- 20 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 59):

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTnG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
25  201 GGATTTGGAT GCCGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

- 30 This corresponds to the amino acid sequence (SEQ ID NO: 60; ORF13):

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLN AGQYVEILRH TGGNRYEVXY RGTWXQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

- 35 Further sequence analysis elaborated the DNA sequence slightly (SEQ ID NO: 61):

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGTTGT
51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTnG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
40  201 GGATTTGGAT GCCGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
251 ACCGTTACGA AGTTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

This corresponds to the amino acid sequence (SEQ ID NO: 62; ORF13-1):

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDL D AGQYVEILRH TGGNRYEVFY RGTHWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 (SEQ ID NO: 60) shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) (SEQ ID NO: 64) from strain A of *N. meningitidis*:

```

10      10      20      30      40      50
orf13.pep      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
orf13a      MTVWFVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
           10      20      30      40      50      60

15      60      70      80      90      100     110
orf13.pep      VHAKTAVRKVETDSYQDL DAGQYVEILRH TGGNRYEVXYRGTXWQAQNTGQEELEPGTRA
orf13a      VHAKTAVGK VETDSYQDL DAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
           70      80      90      100     110     120

20      120
orf13.pep      LIVRKEGNLLIITHPX
orf13a      LIVRKEGNLLIIAKPX
           130

```

The complete length ORF13a nucleotide sequence (SEQ ID NO: 63) is:

```

1  ATGACTGTAT GGT TTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTG GCGGGTTCGG
101 GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151 GCTCTGCTTT CCGCGCTGGG TATTTGGTTC GTACACGCCA AAACCGCCGT
201 GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
251 CCGAAATCCT CCGGCACGCA GGCGGCAACC GTTACGAAGT TTTTATCGC
301 GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
351 AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
401 AACCTTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 64):

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA LAGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
101 GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*

```

ORF13a (SEQ ID NO: 64) and ORF13-1 (SEQ ID NO: 62) show 94.4% identity in 126 aa overlap

		10	20	30	40	50	60
	orf13a.pep	MTVWFVA	AVAVLII	ELLTGT	VYLLV	VSAA	LAGSGIAYGLTGSTPA
5	orf13-1		AVLII	ELLTGT	VYLLV	VSAA	LAGSGIAYGLTGSTPA
			10	20	30	40	50
		70	80	90	100	110	120
	orf13a.pep	VHAKTAV	GKVETDS	YQDLDA	GQYAEIL	RHAGGN	RYEVFYRGTHWQAQNTG
10	orf13-1	VHAKTAV	RKVETDS	YQDLDA	GQYVEIL	RHTGGN	RYEVFYRGTHWQAQNTG
		60	70	80	90	100	110
		130					
	orf13a.pep	LIVRKEGN	LLIIAKPX				
15	orf13-1	LIVRKEGN	LLIITHPX				
			120				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 (SEQ ID NO: 60) shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) (SEQ ID NO: 66) from *N. gonorrhoeae*:

	orf13	AVLII	ELLTGT	VYLLV	VSAA	LAGSGIAYGLTGSTPA	AVLTXALLSALGIXF	51
20	orf13ng	MTVWFVA	AVAVLII	ELLTGT	VYLLV	VSAA	LAGSGIAYGLTGSTPA	60
	orf13	VHAKTAV	RKVETDS	YQDLDA	GQYVEIL	RHTGGN	RYEVXYRGTXWQAQNTG	111
	orf13ng	VHAKTAV	GKVETDS	YQDLDTG	KYAEIL	RYTGGN	RYEVFYRGTHWQAQNTG	120
25	orf13	LIVRKEGN	LLIITHP	126				
	orf13ng	LIVRKEGN	LLIIANP	135				

The complete length ORF13ng nucleotide sequence (SEQ ID NO: 65) is:

30	1	ATGACTGTAT	GGTTTGTTC	CGCTGTTGCC	GTCTTAATCA	TCGAATTATT
	51	GACGGGAACG	GTTTATCTTT	TGGTTGTCAG	CGCGGCTTTG	GCGGGTTCGG
	101	GCATTGCCTA	CGGGCTGACT	GGCAGCACGC	CTGCCGCCGT	CTTGACCGCC
	151	GCACTGCTTT	CCGCGCTGGG	CATTTGGTTC	GTACATGCCA	AAACCGCCGT
	201	GGGAAAAGTT	GAAACGGATT	CATATCAGGA	TTTGGATACC	GGAAAATATG
35	251	CCGAAATCCT	CCGATACACA	GGCGGCAACC	GTTACGAAGT	TTTTTATCGC
	301	GGTACGCACT	GGCAGGCGCA	AAATACGGGG	CAGGAAGTGT	TTGAACCGGG
	351	AACGCGCGCC	CTCATCGTCC	GCAAAGAAGG	TAACCTTCTT	ATCATCGCAA
	401	ACCCTTAA				

40 This encodes a protein having amino acid sequence (SEQ ID NO: 66):

1	MTVWFVA	AVA	VLI	ELLTGT	VYLLV	VSAA	AGSGIAYGLT	GSTPA	AVLTA
51	ALLSALGIWF	VHAKTAV	GKV	ETDSYQDLDT	GKYAEILRYT	GGNRYEVFYR			
101	GTHWQAQNTG	QEVFEPGTRA	LIVRKEGNLL	IIANP*					

ORF13ng (SEQ ID NO: 66) shows 91.3% identity in 126 aa overlap with ORF13-1 (SEQ ID NO: 62):

5	orf13-1.pep	1020304050	AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF
	orf13ng	102030405060	MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF
10	orf13-1.pep	60708090100110	VHAKTAVRKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
	orf13ng	708090100110120	VHAKTAVGKVVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA
15	orf13-1.pep	120	LIVRKEGNLLIITHPX
	orf13ng	130	LIVRKEGNLLIIANPX

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 (SEQ ID NO: 60) and ORF13ng (SEQ ID NO: 66) are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 67):

25	1	ATGTWTGATT	TCGGTTTTrGG	CGArCTGGTT	TTTGTCTGGCA	TTATCGCCCT
	51	GATwGtCCTC	GGCCCCGAAC	GCsTGCCCGA	GGCCGCCCGC	AyCGCCGGAC
30	101	GGcTCATCGG	CAGGCTGCAA	CGCTTTGTCTG	GcAGCGTCAA	ACAGGAATTT
	151	GACACTCAAA	TCGAACTGGA	AGAACTGAGG	AAGGCAAAGC	AGGAATTTGA
	201	AGCTGCCGcC	GCTCAGGTTC	GAGACAGCCT	CAAAGAAACC	GGTACGGATA
	251	TGGAAGGCAA	TCTGCACGAC	ATTTCCGACG	GTCTGAAGCC	TTGGGAAAAA
	301	CTGCCCCGAAC	AGCGGACACC	TGCCGATTTC	GGTGTCTGATG	AAAACGGCAA
	351	TCCGCT.TCC	CGATGCGGCA	AACACCCTAT	CAGACGGCAT	TTCCGACGTT
	401	ATGCCGTC..				

This corresponds to the amino acid sequence (SEQ ID NO: 68; ORF2):

35	1	MXDFLGLGELV	FVGIIALIVL	GPERXPEAAR	XAGRLIGRLQ	RFVGSVKQEF
	51	DTQIELEELR	KAKQEFEEAA	AQVRDSLKET	GTDMEGNLHD	ISDGLKPWEK
	101	LPEQRTPADF	GVDENGNPXS	RCGKHPIRRH	FRRYAV..	

Further work revealed the complete nucleotide sequence (SEQ ID NO: 69):

40	1	ATGTTTGATT	TCGGTTTGGG	CGAGCTGGTT	TTTGTCTGGCA	TTATCGCCCT
----	---	------------	------------	------------	-------------	------------

-120-

51 GATTGTCCTC GGCCCCGAAC GCCTGCCCCGA GGCCGCCCGC ACCGCCGGAC  
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT  
 151 GACACGCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA  
 201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA  
 251 TGGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA  
 301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA  
 351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA  
 401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCCTTG GGACAGCGGG  
 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG  
 501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG  
 551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT  
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAA CGCGATTTTC GTCCGAAACA  
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

15 This corresponds to the amino acid sequence (SEQ ID NO: 70; ORF2-1):

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF  
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNNPLP DAANTLSDGI SDVMPSESY ASAETLGDSG  
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHHT  
 201 SLRKQAISRK RDRPKHRAK PKLRVRKS\*

Further work identified the corresponding gene in strain A of *N.meningitidis* (SEQ ID NO: 71):

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCCGCA TTATCGCCCT  
 51 GATTGTCCTC GGCCCCGAAC GCCTGCCCCGA GGCCGCCCGC ACCGCCGGAC  
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT  
 151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA  
 201 AGCTGCCGCT GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA  
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA  
 301 CTGCCCGAAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA  
 351 TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA  
 401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCCTTG GGACAGCGGG  
 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG  
 501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG  
 551 AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT  
 601 TCGTTCGTA AACAGGCAAT AAGCCGCAA CGCGATTTGC GTCCATAATC  
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This encodes a protein having amino acid sequence (SEQ ID NO: 72; ORF2a):

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF  
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNNPP DAANTLLDGI SDVMPSESY ASAETLGDSG  
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHHT  
 201 SLRKQAISRK RDLRPKSRAK PKLRVRKS\*

45 The originally-identified partial strain B sequence (ORF2) (SEQ ID NO: 68) shows 97.5% identity over a 118aa overlap with ORF2a (SEQ ID NO: 72):

50 orf2.pep      10      20      30      40      50      60  
 MXDFGLGELVFVGIIALIVLGPXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR  
 orf2a      MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR  
                  10      20      30      40      50      60

		70	80	90	100	110	120					
	orf2.pep	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS										
5	orf2a	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP										
		70	80	90	100	110	120					
		130										
	orf2.pep	RCGKHPIRRHFRRYAV										
10	orf2a	DAANTLLDGISDVMPSEERSYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAAPVV										
		130	140	150	160	170	180					

The complete strain B sequence (ORF2-1) (SEQ ID NO: 70) and ORF2a (SEQ ID NO: 72) show 98.2% identity in 228 aa overlap:

15	orf2a.pep	MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
	orf2-1	MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
	orf2a.pep	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP	120
	orf2-1	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP	120
20	orf2a.pep	DAANTLLDGISDVMPSEERSYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAAPVV	180
	orf2-1	DAANTLLDGISDVMPSEERSYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAAPVV	180
	orf2a.pep	QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAPKPLRVKRSX	229
25	orf2-1	QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAPKPLRVKRSX	229

Further work identified a partial DNA sequence (SEQ ID NO: 73) in *N.gonorrhoeae* encoding the following amino acid sequence (SEQ ID NO: 74; ORF2ng):

	1	MFDFGLGELI	FVGIIALIVL	GPRLPEAAR	TAGRLIGRLQ	RFVGSVKQEL
	51	DTQIELEELR	KVKQAFEEAA	AQVRDSLKET	DTDMQNSLHD	ISDGLKPWEK
30	101	LPEQRTPADF	GVDEKGNLSL	RYGKHIRRH	FRYAV*	

Further work identified the complete gonococcal gene sequence (SEQ ID NO: 75):

	1	ATGTTTGATT	TCGGTTTGGG	CGAGCTGATT	TTTGTGCGCA	TTATCGCCCT
	51	GATTGTCCTT	GGTCCAGAAC	GCCTGCCCCG	AGCCGCCCCG	ACTGCCGGAC
35	101	GGCTTATCGG	CAGGCTGCAA	CGCTTTGTAG	GAAGCGTCAA	ACAAGAACTT
	151	GACACTCAAA	TGCAACTGGA	AGAGCTGAGG	AAGGTCAAGC	AGGCATTCTGA
	201	AGCTGCCGCC	GCTCAGGTTC	GAGACAGCCT	CAAAGAAACC	GATACGGATA
	251	TGCAGAACAG	TCTGCACGAC	ATTTCCGACG	GTCTGAAGCC	TTGGGAAAAA
	301	CTGCCCCAAC	AGCGCACGCC	tgccgatttc	gGTGTGATg	AAAacggcaa
40	351	tcccccttccc	gATACGGCAA	ACACCGTATC	AGACGGCATT	TCCGACGTTA
	401	TGCCGTCTGA	ACGTTCCGAT	ACTtccgcCG	AAACCCTTGG	GGACGACAGG
	451	CAAACCGGCA	GTACAGCCGA	ACCTGCGGAA	ACCGACAAAG	ACCGCGCATG
	501	GCGGGAATAC	CTGactgctt	ctgcgcgcgc	acctgtcgta	Cagagggccg
	551	tcgaagtcag	ctaTATCGAT	ACTGCTGTTG	AAacgcctgT	tccgcaCacc
45	601	acttccctgc	gcaAACAGGC	AATAAACCGC	AAACGCGATT	TttgtccgaA
	651	ACACCGCGCC	aaACCGAAat	tgcgcgctcCG	TAAATCATAA	

This encodes a protein having the amino acid sequence (SEQ ID NO: 76; ORF2ng-1):

```

1  MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTMQNSLHD ISDGLKPWEK
101 LPEQRTPADF GVDENGPNLP DTANTVSDGI SDVMPSESD TSAETLGDDR
151 QTGSTAEPAE TDKDRAWREY LTASAAPVV QRAVEVSYID TAVETPVPH
201 TSLRKQAINR KRDFCPKHRA KPKLRVRKS*

```

The originally-identified partial strain B sequence (ORF2) (SEQ ID NO: 68) shows 87.5% identity over a 136aa overlap with ORF2ng (SEQ ID NO: 74):

```

orf2.pep      MXDFGLGELVFVGIIALIVLGPXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng        MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60

orf2.pep      KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNXS 120
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng        KVKQAFEAAAQVRDSLKETDTMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLSP 120

orf2.pep      RCGKHPIRRHFRRYAV 136
| | | | | | | | | |
orf2ng        RYGKHRIRRHFRYAV 136

```

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) (SEQ ID NO: 70 & SEQ ID NO: 76) show 91.7% identity in 229 aa overlap:

```

10      20      30      40      50      60
orf2-1.pep  MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
25 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1    MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
10      20      30      40      50      60

70      80      90      100     110     120
orf2-1.pep  KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNLP
30 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1    KVKQAFEAAAQVRDSLKETDTMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGPNLP
70      80      90      100     110     120

130     140     150     160     170     180
orf2-1.pep  DAANTLSDGISDVMPSESYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAPVV
35 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1    DTANTVSDGISDVMPSESDTSAETLGDDRQTGSTAEPATDKDRAWREYLTASAAPVV
130     140     150     160     170     180

190     200     210     220     229
orf2-1.pep  Q-TVEVSYIDTAVETPVPHHTSLRKQAISRKRDFRPHRAKPKLRVRKSX
40 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf2ng-1    QRAVEVSYIDTAVETPVPHHTSLRKQAINRKDFCPKHRAKPKLRVRKSX
190     200     210     220     230

```



Computer analysis of these amino acid sequences indicates a transmembrane region (underlined), and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein (SEQ ID NO: 1118) of *E.coli*:

```

5      gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
      Score = 56.6 bits (134), Expect = 1e-07
      Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

      Query: 1  MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
                MFD G  EL+ V II L+VLGP+RLP A +T      I  L+      +V+ EL  +++L+E  +
      Sbjct: 1  MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60

10     Query: 61  -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
                +K+  +A+  +   LK +  +++  +
      Sbjct: 61  DSLKKVEKASLTNLTPELKASMDEL RQA 88

```

Based on this analysis, it was predicted that ORF2 (SEQ ID NO: 68), ORF2a (SEQ ID NO: 72) and ORF2ng (SEQ ID NO: 74) are likely to be membrane proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (SEQ ID NO: 70) (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis (Figure 3D). These experiments confirm that ORF37-1 (SEQ ID NO: 4) is a surface-exposed protein, and that it is a useful immunogen.

## 25 Example 10

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 77):

```

30      1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
      51  CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
      101 TTgCGGTCTGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
      151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
      201 CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
      251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
      301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
      351 GTTTGACAGG TTTAACCCTT TCTTTATCTA CACTTAATGC CCCTGCACTC
      401 TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
      451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC

```

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCTGCGC  
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTATTAA  
 601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

5 This corresponds to the amino acid sequence (SEQ ID NO: 78; ORF15):

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFVEQEL VAASARAANK  
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT  
 101 DYTYPRIYETT AETTSGLTGT LTTSLSTLNA PALSRQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVSP ANADTDVFIN  
 201 IDVFGTIRNR TEM..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 79):

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT  
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAA  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA  
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC  
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG  
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT  
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT  
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG  
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG  
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAAC  
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA  
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTT GCAGTAGACA  
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT  
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA  
 801 AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTTCTCC GATATCCGAC  
 851 CATAAGGCAA TCATACGGGT AACTCCGCC CATCCGTAGA GGCTGATAAC  
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA  
 951 AGGACACCT TGA

This corresponds to the amino acid sequence (SEQ ID NO: 80; ORF15-1):

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAANK  
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT  
 101 DYTYPRIYETT AETTSGLTGT LTTSLSTLNA PALSRQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVSP ANADTDVFIN  
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA  
 251 AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIRPYGNHTG NSAPSVADN  
 301 SHEGYGSDE VVRQHRQGP \*

Further work identified the corresponding gene in strain A of *N.meningitidis* (SEQ ID NO: 81):

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAACGCT  
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAA  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA  
 251 TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC  
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG  
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT  
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT

5  
10

```

451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
551 GCATAGACGT TGTCTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC
601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651 TGCCGAAACA CTGAAAGCCC AAACAAAAC TGAATATTTT GCAGTAGACA
701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
801 AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTCTCTC GATATCCAAC
851 CATACGCAAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA
951 AGGGCAACCT TGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 82; ORF15a):

15  
20

```

1  MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAANK
51  DMDLQALHGR KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRYETT AETTSGLTGT LTSLSTLNA PALSRQSDG SGSSSLGLN
151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVVSF ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPYKVSQGIK PTEGLMVDFF DIQPYGNHMG NSAPSVEADN
301 SHEGYGYSDE AVRRHRQGPQ *

```

The originally-identified partial strain B sequence (ORF15) (SEQ ID NO: 78) shows 98.1% identity over a 213aa overlap with ORF15a (SEQ ID NO: 82):

25  
30  
35  
40

```

      10      20      30      40      50      60
orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR
orf15a     MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR
      10      20      30      40      50      60

      70      80      90     100     110     120
orf15.pep  KVALYIATMGDQSGSGLTGGRYSIDAXXXGEYINSPAVRTDYTYPRYETTAETTSGLTGT
orf15a     KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTGT
      70      80      90     100     110     120

      130     140     150     160     170     180
orf15.pep  LTSLSTLNAFALSRQSDGSGSSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
orf15a     LTSLSTLNAFALSRQSDGSGSSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
      130     140     150     160     170     180

      190     200     210
orf15.pep  FLRGIDVVSFANADTDVFINIDVFGTIRNRTEM
orf15a     FLRGIDVVSFANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
      190     200     210     220     230     240

```

45 The complete strain B sequence (ORF15-1) (SEQ ID NO: 80) and ORF15a (SEQ ID NO: 82) show 98.8% identity in 320 aa overlap:

```

      10      20      30      40      50      60
orf15a.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR

```

5	orf15-1	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR	10	20	30	40	50	60
	orf15a.pep	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG	70	80	90	100	110	120
10	orf15-1	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG	70	80	90	100	110	120
	orf15a.pep	LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF	130	140	150	160	170	180
15	orf15-1	LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF	130	140	150	160	170	180
	orf15a.pep	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL	190	200	210	220	230	240
20	orf15-1	FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL	190	200	210	220	230	240
	orf15a.pep	IKPKTNAFEAAAYKENYALWMGPYKVS KG I KPTEGLMVD FSDIQPYGNHMGNSAPSVEADN	250	260	270	280	290	300
25	orf15-1	IKPKTNAFEAAAYKENYALWMGPYKVS KG I KPTEGLMVD FSDIRPYGNHTGNSAPSVEADN	250	260	270	280	290	300
	orf15a.pep	SHEGYGYSDEAVRRHRQGQPX	310	320				
	orf15-1	SHEGYGYSDEVVRQHRQGQPX	310	320				

Further work identified the corresponding gene in *N.gonorrhoeae* (SEQ ID NO: 83):

30	1	ATGCGGGCAC	GGCTGCTGAT	ACCTATTCTT	TTTTCAGTTT	TTATTTTATC
	51	CGCCTGCGGG	ACACTGACAG	GTATTCCATC	GCATGGCGGA	GGCAAACGCT
35	101	TCGCGGTCTGA	ACAAGA A CTT	GTGGCCGCTT	CTGCCAGAGC	TGCCGTTAAA
	151	GACATGGATT	TACAGGCATT	ACACGGACGA	AAAGTTGCAT	TGTACATTGC
40	201	AACTATGGGC	GACCAAGGTT	CAGGCAGTTT	GACAGGGGGT	CGCTACTCCA
	251	TTGATGCACT	GATTCGCGGC	GAATACATAA	ACAGCCCTGC	CGTCCGCACC
45	301	GATTACACCT	ATCCGCGTTA	CGAAACCACC	GCTGAAACAA	CATCAGGCGG
	351	TTTGACGGGT	TTAACCACTT	CTTTATCTAC	ACTTAATGCC	CCTGCACTCT
50	401	CGCGCACCCA	ATCAGACGGT	AGCGGAAGTA	GGAGCAGTCT	GGGCTTAAAT
	451	ATTGGCGGGA	TGGGGGATTA	TCGAAATGAA	ACCTTGACGA	CCAACCCGCG
55	501	CGACACTGCC	TTTCTTTCCC	ACTTGGTGCA	GACCGTATTT	TTCCTGCGCG
	551	GCATAGACGT	TGTTTCTCCT	GCCAATGCCG	ATACAGATGT	GTTTATTAAC
60	601	ATCGACGTAT	TCGGAACGAT	ACGCAACAGA	ACCGAAATGC	ACCTATACAA
	651	TGCCGAAACA	CTGAAAGCCC	AAACAAA A CT	GGAATATTTC	GCAGTAGACA
65	701	GAACCAATAA	AAAATTGCTC	ATCAAACCCA	AAACCAATGC	GTTTGAAGCT
	751	GCCTATAAAG	AAAATTACGC	ATTGTGGATG	GGGCCGTATA	AAGTAAGCAA
70	801	AGGAATCAAA	CCGACGGAAG	GATTGATGGT	CGATTTCTCC	GATATCCAAC
	851	CATACGGCAA	TCATACGGGT	AACTCCGCC	CATCCGTAGA	GGCTGATAAC
75	901	AGTCATGAGG	GGTATGGATA	CAGCGATGAA	GCAGTGC GAC	AACATAGACA
	951	AGGGCAACCT	TGA			

This encodes a protein having amino acid sequence (SEQ ID NO: 84; ORF15ng):

5  
1 MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK  
51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT  
101 DYTYPREYET AETTSGGLTG LTTSLSTLNA PALSRQSDG SGRSSLGLN  
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN  
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA  
251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN  
301 SHEGYGSDE AVRQHRQQP \*

10 The originally-identified partial strain B sequence (ORF15) (SEQ ID NO: 78) shows 97.2% identity over a 213aa overlap with ORF15ng (SEQ ID NO: 84):

15  
orf15.pep MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKMDLQALHGR 60  
|:|||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf15ng MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKMDLQALHGR 60  
  
orf15.pep KVALYIATMGDQSGSLTGGGRYSIDAXXXGEYINSPAVRTDYTPRYETTAETTSGGLTG 120  
|||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf15ng KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG 120  
  
orf15.pep LTTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF 180  
|||||||||||||||||||||||||:|||||||||||||||||||||||||  
orf15ng LTTSLSTLNAPALSRQSDGSGSRSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF 180  
  
20 orf15.pep FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM 213  
|||||||||||||||||||||  
orf15ng FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL 240

25 The complete strain B sequence (ORF15-1) (SEQ ID NO: 80) and ORF15ng (SEQ ID NO: 84) show 98.8% identity in 320 aa overlap:

30  
orf15-1.pep 10 20 30 40 50 60  
MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKMDLQALHGR  
|:|||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf15ng MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKMDLQALHGR  
10 20 30 40 50 60  
  
orf15-1.pep 70 80 90 100 110 120  
KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG  
|||||||||||||||||||||||||||||||||||||||||||||||||||||  
orf15ng KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGGLTG  
70 80 90 100 110 120  
  
orf15-1.pep 130 140 150 160 170 180  
LTTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF  
|||||||||||||||||||||:|||||||||||||||||||||  
orf15ng LTTSLSTLNAPALSRQSDGSGSRSSLGLNIGMGDYRNETLTTNPRDTAFLSHLVQTVF  
130 140 150 160 170 180  
  
orf15-1.pep 190 200 210 220 230 240  
FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL  
|||||||||||||||||||||  
orf15ng FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL  
190 200 210 220 230 240  
  
45 250 260 270 280 290 300

orf15-1.pep	IKPKTNAFEAAAYKENYALWMGPYKVS	KG	IKPTEGLMVD	FSDIRPYGNHTGNSAPSVEADN
orf15ng	IKPKTNAFEAAAYKENYALWMGPYKVS	KG	IKPTEGLMVD	FSDIQPYGNHTGNSAPSVEADN
	250	260	270	280 290 300
orf15-1.pep	SHEGYGYSDE	VVRQHRQGQPX		
orf15ng	SHEGYGYSDE	AVRQHRQGQPX		
	310	320		

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (SEQ ID NO: 80) (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 11

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 85):

```

1  ..GG.CAGCACA AAAACAGGC GGTGAACGG AAAAACCGTA TTTACGATGA
51  TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATTCTC CGCAAAATAT
101 ATCCCCGCGT TCGGGCTTCA AATTTCTTC ATCCTGTTTT TAACCGCCGT
151 CGCATTCAAA AACTGCATA CCGACCCTCA GACGGCATCC CGCCCGCTGC
201 CCGGACTGCC CrGACTGACT GCGGTTTCCA CACTGTTCGG CACAATGTCTG
251 AGCTGGGTCTG GCATAGGCGG CGGTTCACTT TCCGTCCCCT TCTTAATCCA
301 CTGCGGCTTC CCCGCCATA AAGCCATCGG CACATCATCC GGCCTTGCTT
351 GGCCGATTGC ACTTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT
401 ATTGCAGGAT TGCCCGAAGG GTCCTGTTTACC TGCCCGCCGT
451 CGCCGTCCTC AGCGCGGCAA CCATTGCCTT TGCCCGGCTC GGTGTCAAAA
501 CCGCCACAA ACTTCTTCT GCAAACCTCA AAAAATC.TT CGGCATTATG
551 TTGCTTTTGA TTGCCGGAAT AATGCTGTAC AACCTGCTTT AA

```

This corresponds to the amino acid sequence (SEQ ID NO: 86; ORF17):

```

1  ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
51  AFKTLHTDPQ TASRPLPGLP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH
101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPVAV
151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 87):

```

1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGC GGC
51  AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
101 CTGTTCGTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC
151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
201 CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
251 CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351 GTTTTTAAAC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
401 CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCACACTG
451 TTCGGCACAA TGTCGAGCTG GGTCGGCATA GGCGGCGGTT CACTTTCCGT
501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601 CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
651 TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAAAA
751 Tc.TTCGCA TTATGTTGCT TTTGATTGCC GAAAAATGC TGTACAACCT
801 GCTTTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 88; ORF17-1):

```

1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51  AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTVFTMMP GMIFGVFTGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL
151 FGTMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLVKTA HKLSSAKLKK
251 XFGIMLLLIA GKMLYNLL*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070) (SEQ ID NO: 1119)

ORF17 (SEQ ID NO: 86) and HI0902 proteins (SEQ ID NO: 1119) show 28% aa identity in 192 aa overlap:

```

ORF17  3  HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
          HK  +  +  V  +  P  ++  VF  G  F  +  +IF  +++L  ++  D
HI0902 72  HKLGNIVWQAVRILAPVIMLSVFICGLFGRDLREISAKIFACLVVYLATKMLVLSIKKD- 130

ORF17  60  QTASRPLPGLPXLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPI 119
          Q  ++  L  L  +  L  G  SS  GIGGG  VPFL  G  +AIG+S+  +
HI0902 131  QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189

ORF17  120  ALSGAISYLLNGLNIAGLPEGSLGFLYLPVAVLSAATIAFAPLVGXXXXXXXXXXXXXXX 179
          +SG  S++++G  +PE  SLG++YLPVAV  ++A  +  +  LG
HI0902 190  GISGMFSFIVSGWGNPLMPEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKG 249

```

ORF17 180 FGIMLLLIAGKM 191  
F + L+++A M  
HI0902 250 FALFLIVVAINM 261

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 5 ORF17 (SEQ ID NO: 86) shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) (SEQ ID NO: 90) from strain A of *N. meningitidis*:

	orf17.pep	GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS	10	20	30			
10	orf17a	QGLAQHPYAQH <sup>LAV</sup> GTSTFAVMVF <sup>TAF</sup> S <sup>SML</sup> LGQHKKQAVDWKTVFTMMPGMVFGVFAGALS	50	60	70	80	90	100
	orf17.pep	AKYIPAFGLQIFFILFLTA <sup>VAF</sup> KTLHTDPQTASRPLPGLPXLTA <sup>VSTL</sup> FGMTSSWVGIGG	40	50	60	70	80	90
15	orf17a	AKYIPAFGLQIFFILFLTA <sup>VAF</sup> KTLHTDPQTASRPLPGLPGLTAVSTLFGMTSSWVGIGG	110	120	130	140	150	160
	orf17.pep	GSLSPVFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGS LGFLYLPA <sup>V</sup>	100	110	120	130	140	150
20	orf17a	GSLSPVFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGS LGFLYLPA <sup>V</sup>	170	180	190	200	210	220
	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX	160	170	180	190		
25	orf17a	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX	230	240	250	260		

The complete length ORF17a nucleotide sequence (SEQ ID NO: 89) is:

30	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGCGGC
	51	AGGTTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCCTCACCGC
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GA CTGGAAAA
35	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCACTCT
	351	GTTTTTAAAC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	TCCTCAGACG
	401	CATCCC GCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCCACTG
	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CAC TTTCCGT
40	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAGGGTCAC	TGGGCTTCCT
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAAACTTT	CTTCTGCCAA	ACTCAAAAAA
45	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTTAA				

This encodes a protein having amino acid sequence (SEQ ID NO: 90):



5  
 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY  
 51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTVFTMMP GMVFGVFAGA  
 101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL  
 151 FGTMSWVG I GGGSLVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
 201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK  
 251 SFGIMLLLIA GKMLYNLL\*

ORF17a (SEQ ID NO: 90) and ORF17-1 (SEQ ID NO: 88) show 98.9% identity in 268 aa overlap:

10	orf17a.pep	10 20 30 40 50 60	MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVL DLQGLAQHPYAQHLAVGTSF
	orf17-1	10 20 30 40 50 60	MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVL DLQGLAQHPYAQHLAVGTSF
15	orf17a.pep	70 80 90 100 110 120	AVMVFTAFSSMLGQHKQAVDWKTVFTMMPGMVFGVFAGALSAKYIPAFGLQIFFILFLT
	orf17-1	70 80 90 100 110 120	AVMVFTAFSSMLGQHKQAVDWKTVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
20	orf17a.pep	130 140 150 160 170 180	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLVPFLIHCGFPAHKA
	orf17-1	130 140 150 160 170 180	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLVPFLIHCGFPAHKA
25	orf17a.pep	190 200 210 220 230 240	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP AVAVLSAATIAFAPLGVKTA
	orf17-1	190 200 210 220 230 240	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP AVAVLSAATIAFAPLGVKTA
30	orf17a.pep	250 260 269	HKLSSAKLKKSFIMLLLIAGKMLYNLLX
	orf17-1	250 260	HKLSSAKLKKXFGIMLLLIAGKMLYNLLX

### Homology with a predicted ORF from *N.gonorrhoeae*

35 ORF17 (SEQ ID NO: 86) shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) (SEQ ID NO: 92) from *N. gonorrhoeae*:

	orf17.pep	GQHKQAVNGKTVFTMMPGMIFGVFTGAFS	30
	orf17ng	QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKQAVDWKTI FAMMPGMIFGVFAGALS	102
40	orf17.pep	AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSSWVGIGG	90
	orf17ng	AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGG	162
	orf17.pep	GSLSPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP AV	150
45	orf17ng	GSLSPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLP AV	202

orf17.pep AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLL 196  
 orf17ng AVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKMLYNLL 268

An ORF17ng nucleotide sequence (SEQ ID NO: 91) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 92):

5 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY  
 51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA  
 101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL  
 151 FGAMSSWVG I GGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
 10 201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE  
 251 SFGIMLLLIA GKMLYNLL\*

Further work revealed the complete gonococcal DNA sequence (SEQ ID NO: 93):

15 1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCcgtag gcAGTGCGGC  
 51 AGGTTTTATT GCCGGCCTGT Tcgggtgtagg cggcgGTACG CTGATTGTCC  
 101 CTGTCGTTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC  
 151 GCGCAACACC TCGCCGTCGG CACaTccttc gcCGTCATGG TCTTCACCGC  
 201 CTTTTCCAGT ATGTTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA  
 251 CCATATTGCG GATGATGCCG GGTATGATAT TCGGCGTATT CGCTGGCGCA  
 20 301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT  
 351 GTTTTAAACC GCCGTCGCAT TCAAAACACT GCATACCGGT CGTCAGACGG  
 401 CATCCCGCCC GCTGCCCGGG CTGCCCGGAC TGA CTGCGGT TTCCACACTG  
 451 TTCGGCGCAA TGTCGAGCTG GGTCCGCATA GGCGCGGGT CACTTTCCGT  
 501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT  
 25 551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG  
 601 GTCAACGGTC TGAATATTGC AGGATTGCCC GAAGGGTCGC TGGGCTTCCT  
 651 TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCC  
 701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAGAA  
 751 TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT  
 30 801 GCTTTAA

This corresponds to the amino acid sequence (SEQ ID NO: 94; ORF17ng-1):

35 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY  
 51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA  
 101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL  
 151 FGAMSSWVG I GGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
 201 VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE  
 251 SFGIMLLLIA GKMLYNLL\*

40 ORF17ng-1 (SEQ ID NO: 94) and ORF17-1 (SEQ ID NO: 88) show 96.6% identity in 268 aa overlap:

45 orf17-1.pep MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF  
 orf17ng-1 MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVLDLQGLAQHPYAQHLAVGTSF  
 50 orf17-1.pep AVMVFTAFSSMLGQHKQAVDWKTIVFTMMPGMIFGVFTGALS AKYIPAFGLQIFFILFLT  
 orf17ng-1 AVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVFAGALS AKYIPAFGLQIFFILFLT

		130	140	150	160	170	180
	orf17-1.pep	AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA					
5	orf17ng-1	AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGGGSLSVPFLIHCGFPAHKA					
		130	140	150	160	170	180
	orf17-1.pep	IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA					
10	orf17ng-1	IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA					
		190	200	210	220	230	240
	orf17-1.pep	HKLSSAKLKKXFGIMLLLIAGKMLYNLLX					
15	orf17ng-1	HKLSSAKLKESFGIMLLLIAGKMLYNLLX					
		250	260	269			

In addition, ORF17ng-1 (SEQ ID NO: 94) shows significant homology with a hypothetical *H.influenzae* protein (SEQ ID NO: 1119):

20	sp P44070 Y902_HAEIN_HYPOTHETICAL_PROTEIN_HI0902_pir G64015_hypothetical_protein_HI0902 - Haemophilus influenzae (strain Rd KW20) gi 1573922 (U32772) H. influenzae predicted coding region HI0902 [Haemophilus influenzae]Length = 264 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 15/43 (34%), Positives = 23/43 (53%)	
25	Query: 55 AVGTSFAVMVFTAFSSMLGQHKKQAVDWKTI FAMMPGMIFGVF 97 A+GTSFA +V T S HK + W+ + + P ++ VF Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94  Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 44/114 (38%), Positives = 65/114 (57%)	
30	Query: 150 LFGAMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209 L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G + Sbjct: 148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207  Query: 210 PEGSLGFLYLPAAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263 PE SLG++YLPV ++A + + LG KL + LK+ F + L+++A M Sbjct: 208 PEYSLGYIYLPVVGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261	

35 This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 12

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 95):

-134-

5  
1 ..GGAAACGGAT GGCAGGCAGA CCCCGAACAT CCGCTGCTCG GGCTTTTTGC  
51 CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTGCGAATA TGTGCGTTGG  
101 TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTTGTGTT TGCGGCACTG  
151 CTCAAACTTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT  
201 GCTGATGGCG GTTGCCTATG TCCACCGCTG CCGTATAGAC CGGCAGCCGC  
251 CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGGCAGCG  
301 TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGAAGATA  
351 A

10 This corresponds to the amino acid sequence (SEQ ID NO: 96; ORF18):

1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL  
51 LKLYALKPVY WFLVLFVLMVA VAYVHRCGID RQPPSTFGGS QLRLGGLTAA  
101 LMQVSVLVLL LSEIGR\*

15 Further work revealed the complete nucleotide sequence (SEQ ID NO: 97):

20  
25  
1 ATGATTTTGC TGCATTTGGA TTTTTTGTCT GCCTTACTGT ATGCGGCGGT  
51 TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA  
101 GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC  
151 GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTAA  
201 CCTGACTTTG GGCAGCATAT TTTTTTTCAT CGGGCATTGG AACC GGAAAA  
251 CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGGCTT  
301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTGC GAATATGTGC  
351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTC GTGTTTGCAG  
401 CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG  
451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA  
501 GCCGCCGTCA ACGTTCGGCG GCTCGCAGCT GCGACTCGGC GGGTTGACGG  
551 CAGCGTTGAT GCAGGTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA  
601 AGATAA

This corresponds to the amino acid sequence (SEQ ID NO: 98; ORF18-1):

30  
35  
1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP  
51 GIWGMTRAAP LFIPHFYLT LGSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL  
101 FAVSNVSMTL AFVIGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ  
151 FVLMAVAVVH RCGIDRQPPS TFGGSQRLRG GLTAALMQVS VLVL LLLSEIG  
201 R\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 (SEQ ID NO: 96) shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) (SEQ ID NO: 100) from strain A of *N. meningitidis*:

40  
orf18.pep  
orf18a  
60 70 80 90 100 110  
GNGWQADPEHPLLGLFAVSNVSMTLAFVGI  
|||

5

	40	50	60	70	80	90
orf18.pep	<u>CALVHYCFSGTQVFVFAALLKLYALKPVIYFVLQFVLMAYAVVHRCGIDRQPPSTFGGS</u>					
orf18a	CALVHYCFSXTQVFVFAALLKLYALKPVIYFVLQFVLMAYAVVHRCGIDRQPPSTFGGS					
	120	130	140	150	160	170

10

	100	110	
orf18.pep	<u>QLRLGGLTAALMQSVLVLLSEIGRX</u>		
orf18a	QLRLGGLTAALMQSVLVLLSEIGRX		
	180	190	200

The complete length ORF18a nucleotide sequence (SEQ ID NO: 99) is:

	1	ATGATTTTGC	TGCATTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
15	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCC
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TTGTTTCATC	CCCATTTTTTA
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATTTG	AACCGGAAAA
	251	CGGATGGAAA	CGGATGGCAG	GCAGACCCG	AACATCCTCT	GCTCGGCGTG
20	301	TTTGCCGTA	GTAATGTATC	GATGACGCTT	GCTTTTGTCT	GAATATGTGC
	351	GTTGGTGCAT	TATTGCTTTT	CGNGAACGGT	TCAAGTGTTT	GTGTTTGC GG
	401	CACTGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCAG
	451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
	501	GCCGCCGTCA	ACGTTTCGGC	GNTCGCAGCT	GCGACTCGGC	GGGTTGACGG
25	551	CAGCGTTGAT	GCAGNTCTCG	GTACTGGTGC	TGCTGCTTTC	AGAAATTGGA
	601	AGATAA				

This encodes a protein having amino acid sequence (SEQ ID NO: 100):

30

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP
51	<u>GIWGMTRAAP</u>	<u>LFIPHFYLT</u>	<u>LSIFFFI</u>	<u>GHWN</u>	<u>NRKTDGNGWQ</u>
101	FAVSNVSMTL	AFVGICALVH	YCFSXTQVVF	VFAALLKLYA	LKPVYWFVLQ
151	FVLMAYAVVH	RCGIDRQPPS	TFGGSQRLRG	<u>GLTAALMQXS</u>	<u>VLVLLLSEIG</u>
201	R*				

ORF18a (SEQ ID NO: 100) and ORF18-1 (SEQ ID NO: 98) show 99.0% identity in 201 aa  
35 overlap:

		10	20	30	40	50	60
	orf18a.pep	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFASIMLWLGISVLGAKLMPGIWGMTRAAP					
40	orf18-1	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFASIMLWLGISVLGAKLMPGIWGMTRAAP					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf18a.pep	LFIPHFYLTLGSIFFFIHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH					
45	orf18-1	LFIPHFYLTLGSIFFFIHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf18a.pep	YCFSTXVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQRLRG					
50	orf18-1	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQRLRG					
		130	140	150	160	170	180

```

                                190      200
orf18a.pep  GLTAALMQXSVLVLLLSEIGRX
             |||||
orf18-1     GLTAALMQVSVLVLLLSEIGRX
0           190      200

```

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 (SEQ ID NO: 96) shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) (SEQ ID NO: 102) from *N. gonorrhoeae*:

10	orf18.pep	GNGWQADPEHPLLGLFAVSNVSMTLAFVGI	30
	orf18ng		
	orf18ng	TRAAPLFIPHFYLTLSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI	115
	orf18.pep	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGS	90
	orf18ng		
	orf18ng	CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAYVHRCGIDRQPPSTFGGS	175
15	orf18.pep	QLRLGGLTAALMQVSVLVLLSEIGR	116
	orf18ng	:      :  ::  :	
	orf18ng	QLRLGVLAALMLQVAVTAMLLAEIGR	201

The complete length ORF18ng nucleotide sequence is (SEQ ID NO: 101):

20	1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGt	aTGCGGcggt
	51	tttTctgTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTGCGTT	GTGGCTCGGC	ATCTCGGTTT	TAGGGGTAAA	GCTGATGCCG
	151	GGGATGTGGG	GAATGACCCG	CGCCGCGCCT	TTGTTTCATC	CCCATTTTTTA
	201	CGTGACTTTG	GGCAGCATAT	TTTTTTTTCAT	CGGGTATTGG	AACCGGAAAA
25	251	CAGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
	301	TTTGCCGTCA	GTAATGTATC	GATGACGCTT	GCTTTTGTCTG	GAATATGTGC
	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	TCAAGTGTTT	GTGTTTGCGG
	401	CATTGCTCAA	ACTTTATGCG	CTGAAGCCCG	TTTATTGGTT	CGTGTTCAG
	451	TTTGTATTGA	TGGCGGttgC	CTATGTCCAC	CGCTGCGGTA	TAGACCGGCA
30	501	GCCGCCGTCA	ACGTTCCGGC	GTTCGAGCT	GCGACTCGGC	GTGTTGGCGG
	551	CGATGTTGAT	GCAGGTTGCG	GTAACGGCGA	TGCTGCTTGC	CGAAATCGGC
	601	AGATGA				

This encodes a protein having amino acid sequence (SEQ ID NO: 102):

35           1   MILLHLDFLS   ALLYAAVFLF   LIFRAGMLQW   FWASIALWLG   ISVLGVKLMF  
          51   GMWGMTRAAP   LFIPHFYLT   GSIFFFIGYW   NRKTDGNGWQ   ADPEHPLLGL  
         101   FAVSNVSMTL   AFVGICALVH   YCFSGTVQVF   VFAALLKLYA   LKPVYWFVLQ  
         151   FVLMAVAYVH   RCGIDRQPPS   TFGGSQRLRG   VLAAMLQVA   VTAMLLAEIG  
40           201   R\*

This ORF18ng (SEQ ID NO: 102) protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1 (SEQ ID NO: 98):

orf18-1.pep                    10                    20                    30                    40                    50                    60  
MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP

	orf18ng	 MILLHLDFLSALLYAAVFLFLIFRAGMLQFWASIALWLGISVLGVKLMPGMWGMTRAAP	10	20	30	40	50	60
5	orf18-1.pep	LFIPHFYLTLSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH	70	80	90	100	110	120
	orf18ng	 LFIPHFYLTLSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH	70	80	90	100	110	120
10	orf18-1.pep	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGSQRLRG	130	140	150	160	170	180
	orf18ng	 YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYVYVHRCGIDRQPPSTFGGSQRLRG	130	140	150	160	170	180
15	orf18-1.pep	GLTAALMQVSVLVLLSEIGRX	190	200				
	orf18ng	:      :  ::  :      VLAAMLQVAVTAMLLAEIGRX	190	200				

Based on this analysis, including the presence of several putative transmembrane domains in the  
 20 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 13

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 103):

25	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTN	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTT	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTTGTCGAT
	151	TTGGACAACC	NCNTGACCGG	ACGGCTNAAA	AACATCATCA	CCACCGTCGC
	201	CCTGTTCACC	CTCTCCTCGC	TCACGGCACA	AAGCACCTC	GGCACAGGGC
30	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CTT.CG.CTT	CACCATTTTA
	301	GGCGCGNCG	...			

This corresponds to the amino acid sequence (SEQ ID NO: 104; ORF19):

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
	51	LDNXXTGRLK	NIITTVALFT	LSSSLAQSTL	GTGLPFILAM	TLMTXXFTIL
	101	GAX...				

Further work revealed the complete nucleotide sequence (SEQ ID NO: 105):

40	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTT	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTTGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCA	CCACCGTCGC
	201	CCTGTTCACC	CTCTCCTCGC	TCACGGCACA	AAGCACCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	CACCATTTTA

5  
10  
15  
20  
25  
30  
35

```

301  GCGCGGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
351  CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
401  ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
451  CTGTTCCAAA TCGTCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
501  CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
551  ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
601  AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
651  TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACC GCCC AAAATGCTGC
701  GTTACTACTT TGCGGCCCAA GACATACACG AACGCATCAG CTCCGCCAC
751  GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAAACCG ACATCATCTT
801  CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
851  CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
901  CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
951  CGACAGTCCC GACATCCGCC ACCTGCGCGC CCTTCTCGAC AACCTCGGCA
1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051 AACGACCGCA TGCGCGCAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CGTCTGTGCA AACCAAATC TGGATTGTCA TCGCCAGTAC
1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551 TGCTTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAG TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCGTTTCCGA CATGAGCAGC GAACCCGCAA AATTTCGCCG CAGCCTGCAA
1801 CCCGGCTTTA CCTGTCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCC AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
2001 GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051 ACATCTCTCT CCAACAGCTC CAACTCATCG CCGACAGCT CGAACCTAC
2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
2151  A

```

40 This corresponds to the amino acid sequence (SEQ ID NO: 106; ORF19-1):

45  
50  
55

```

1  MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51  LDNRLTGRLK NIITTVLFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT
401 IVEALNLLNG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQUAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPETEPDDF QTALDTRLGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPNAA*

```

Computer analysis of this amino acid sequence gave the following results:



Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289) (SEQ ID NO: 1120)

ORF19 (SEQ ID NO: 104) and YHFK proteins (SEQ ID NO: 1120) show 45% aa identity in 97 aa overlap:

```

5      orf19  6  LKPLLITSLPVPFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65
      L   +I+++PVF +V AA +W           +MP +LGIIAGGLVDLDN TGRLKN+ T
      YHFK   5  LNAKVISTIPVFIADVIAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRLKNVFFT 64

      orf19  66  VALFTLSSSLTAQSTLGTGLPFILAMTLMTXXFTILGA 102
      + F++SS Q +G + +I+ MT++T FT++GA
10     YHFK   65  LIAFSISSFIVQLHIGKPIQYIVLMTVLTFTFTMIGA 101

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 (SEQ ID NO: 104) shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) (SEQ ID NO: 108) from strain A of *N. meningitidis*:

```

15      orf19.pep  MKTPLLKPLLITSLPVPFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK
      orf19a      MKTPLLKPLLITSLPVPFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
                  10      20      30      40      50      60

20      orf19.pep  NIITTVLFTLSSSLTAQSTLGTGLPFILAMTLMTXXFTILGAX
      orf19a      NIIATVALFTLSSSLVAQSTLGTGLPFILAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY
                  70      80      90      100      110      120

25      orf19a      TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA
                  130      140      150      160      170      180

```

The complete length ORF19a nucleotide sequence (SEQ ID NO: 107) is:

```

1  ATGAAAACCC CACCCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
51 CGCCAGTGTC TTTACCGCCG CCTCCATCGT CTGGCAGCTG GGCGAACCCA
30 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCTGGCGG CCTGGTCGAT
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCG CCACCGTCGC
201 CCTGTTCACC CTCTCCTCAC TTGTGCGGCA AAGCACCCCTC GGCACAGGTT
251 TGCCATT CAT CCTCGCCATG ACCCTGATGA CTTTCGGCTT TACCATCATG
301 GCGCGGTCG GGCTGAAATA CCGCACCTTC GCCTTCGGCG CACTCGCCGT
35 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
401 ACCCCTTAT GATTCTGTGC GGAACCGTAC TGTACAGCAC CGCCATCATC
451 CTGTTCCAAA TCATCCTGCC CCACCGCCCC GTTCAAGAAA ACGTCGCCAA
501 CGCCTACGAA GCACTCGGCA GCTACCTCGA AGCCAAAGCC GACTTTTTCG
551 ATCCCGACGA AGCCGAATGG ATAGGCAACC GCCACATCGA CCTCGCCATG
40 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
701 GCTACTACTT CGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCAC

```

5  
10  
15  
20  
25  
30

```

751  GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAAACACCG ACATCATCTT
801  CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
851  CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
901  CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
951  CGACAATCCC GACATCCGCC ACCTGCGCCG CCTTCTCGAC AACCTCGGCA
1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCGGCAGCCT
1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTTG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTTACCC CCTCCGTCGA AACCAAACCTC TGGATCGTCA TCGCCAGTAC
1401 CACCCCTCTT TTCATGACCC GCACCTACAA ATACAGCTTC TCGACATTTT
1451 TCATCACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG GTTGGACGTA
1501 TACGCCGCCA TGCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551 TGCCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGCGC CTATCTCGAA
1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801 CCCGGCTTTA CCTGTCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCC AAACCGAACC CGAAGACTTT CAGACAGCAC TGGATACACT
2001 GCGCGCGGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGGCAGCT CGAACCCTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151  A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 108):

35  
40  
45

```

1  MKTPPLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51  LDNRLTGR LK NIIATVALFT LSSLVAQSTL GTGLPFILAM TLMTFGFTIM
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI I
151 LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIH RLLEMGGQAC RNTAQALRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSNDNP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETGSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT
401 IVEALN LNLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITI QAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLT LERTAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLK TGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPETEPDDF QTALD TLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

```

ORF19a (SEQ ID NO: 108) and ORF19-1 (SEQ ID NO: 106) show 98.3% identity in 716 aa overlap:

50

```

      10      20      30      40      50      60
orf19a.pep MKTPPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGR LK
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
orf19-1     MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGR LK
      10      20      30      40      50      60

```

		70	80	90	100	110	120
	orf19a.pep	NIIATVALFTLSSSLVAQSTLGTGLPFI	LAMTLM	TGFTIM	GAVGLKYRT	FAFGALAVATY	
	orf19-1	NIITVALFTLSSSLTAQSTLGTGLPFI	LAMTLM	TGFTIL	GAVGLKYRT	FAFGALAVATY	
5		70	80	90	100	110	120
		130	140	150	160	170	180
	orf19a.pep	TTLTYTPETYWLTNPFMILCGTVLYSTAI	ILFQIIL	PHRPVQEN	VANAYEALGSYLEAKA		
	orf19-1	TTLTYTPETYWLTNPFMILCGTVLYSTAI	LLFQIVL	PHRPVQES	VANAYDALGGYLEAKA		
10		130	140	150	160	170	180
		190	200	210	220	230	240
	orf19a.pep	DDFDPDEAEWIGNRHIDLAMSN	TGVITAFNQCRSALFYRLRGK	HRHPRTAKMLRYYFAAQ			
	orf19-1	DDFDPDEAAWIGNRHIDLAMSN	TGVITAFNQCRSALFYRLRGK	HRHPRTAKMLRYYFAAQ			
15		190	200	210	220	230	240
		250	260	270	280	290	300
	orf19a.pep	DIHERISSAHVDYQEMSEKFKNTDII	FRIHRLLEM	QOQACRNTA	QALRASKDYVYSKRLG		
	orf19-1	DIHERISSAHVDYQEMSEKFKNTDII	FRIHRLLEM	QOQACRNTA	QALRASKDYVYSKRLG		
20		250	260	270	280	290	300
		310	320	330	340	350	360
	orf19a.pep	RAIEGCRQSLRLLSDSN	DPDIRHLRRLLDNLGSVDQ	QFRLQHNLQAENDRM	GDTRIA		
	orf19-1	RAIEGCRQSLRLLSDSN	DPDIRHLRRLLDNLGSVDQ	QFRLQHNLQAENDRM	GDTRIA		
25		310	320	330	340	350	360
		370	380	390	400	410	420
	orf19a.pep	ALETGSLKNTWQAIRPOLN	LESGVFRHAVRLSLVVA	ACTIVEALN	NLNGYWILLTALFV		
	orf19-1	ALETSSLKNTWQAIRPOLN	LESGVFRHAVRLSLVVA	ACTIVEALN	NLNGYWILLTALFV		
30		370	380	390	400	410	420
		430	440	450	460	470	480
	orf19a.pep	CQPNYTATKSRVRQRIAG	TVLGVIVGSLVPYFTPS	VETKLWIVIA	STTLFFMTRTYKYSF		
	orf19-1	CQPNYTATKSRVRQRIAG	TVLGVIVGSLVPYFTPS	VETKLWIVIA	STTLFFMTRTYKYSF		
35		430	440	450	460	470	480
		490	500	510	520	530	540
	orf19a.pep	STFFITIQA	LTSLSLAGLDVYAAMP	VRIIDTII	IGASLAWAAVS	YLWPDWKYLT	LERTAAL
	orf19-1	STFFITIQA	LTSLSLAGLDVYAAMP	VRIIDTII	IGASLAWAAVS	YLWPDWKYLT	LERTAAL
40		490	500	510	520	530	540
		550	560	570	580	590	600
	orf19a.pep	AVCSNGAYLEKITERLK	SGETGDDVEYRATRRRA	HEHTAALSSTLS	SDMSSEPAKFADSLQ		
	orf19-1	AVCSNGAYLEKITERLK	SGETGDDVEYRATRRRA	HEHTAALSSTLS	SDMSSEPAKFADSLQ		
45		550	560	570	580	590	600
		610	620	630	640	650	660
	orf19a.pep	PGFTLLKTGYALTGYIS	ALGAYRSEMHEECSPD	FTAQFHLAAEHTA	HIHQHLPETEPDDF		
	orf19-1	PGFTLLKTGYALTGYIS	ALGAYRSEMHEECSPD	FTAQFHLAAEHTA	HIHQHLPETEPDDF		
50		610	620	630	640	650	660

		670	680	690	700	710
orf19a.pep		QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHRQPQNAAX				
orf19-1		QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHRQPQNAAX				
5		670	680	690	700	710

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 (SEQ ID NO: 104) shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) (SEQ ID NO: 110) from *N. gonorrhoeae*:

10	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK	60
	orf19.pep	NIITTVALFTLSSSLTAQSTLGTGLPFILAMTLMTXXTILGAX	103
15	orf19ng	NIITATVALFTLSSSLTAQSTLGTGLPFILAMTLMTFGTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence (SEQ ID NO: 109) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 110):

	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIAGGLVD
	51	LDNRLTGRLK	NIIATVALFT	LSSSLTAQSTL	GTGLPFILAM	TLMTFGFTIL
20	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAI
	151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DFDPEAAW	IGNRHIDLAM
	201	SNTGVITAFN	QCRSALFYRL	RGKHRHPTA	KMLRYFAAQ	DIHERISSAH
	251	VDYQEMSEKF	KNTDIIFRIR	RLLEMQQAC	RNTAQAIRSG	KDYVYSKRLG
	301	RAIEGCRQSL	RLLSGDNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPA
25	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence (SEQ ID NO: 111):

	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
30	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGCGCG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTTACC	CTCTCCTCGC	TCACGCGCGA	AAGCACCTCT	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
	301	GGCGCGGTCT	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
35	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCACAA	TCATCCTGCC	CCACCGCCCC	GTCCAAGAAA	CGCTCGCCAA
	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCCGATGA	GGCAGCCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
40	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCGCATCCGC	CGCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
45	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	tTTACAGCAA	ACGCCTCGGA
	901	CGCGCCATcg	aaggctgCCG	CCAGTCGctg	cgcctCCTTt	cagacggcaA
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGgca
	1001	GCGTcgacca	gcagtTCcgc	caactCCGAC	ACAgcgactC	CCCCGcgaa

5  
10  
15  
20  
25

```

1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCCtcgaaa ccggcagctT
1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC
1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CCTCCGTCGA AACCAAACCTC TGGATTGTCA TCGCCGGTAC
1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TACGCCGCCA TGCCCGTGCG CATCATcgaC ACCATTATCG GCGCATCCCT
1551 TGCCTGGGCG GCGGTGAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
1651 AAAATTGCCG AACGCCTCAA AACCGCGGAA ACCGCGGACG ACATAGAATA
1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801 CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT
2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC
2051 ACATCTCCT CCAACAGCTC CAACTCATCG CcggGCAACT CGAACCCCTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151 A

```

25 This corresponds to the amino acid sequence (SEQ ID NO: 112; ORF19ng-1):

30  
35  
40

```

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI I
151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIR RLLEMQQQAC RNTAQAIRSG KDYVYSKRLG
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA E
351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAAC T
401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSP STFFITI QAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSSGTYLQ
551 KIAERLKTGE TGDDIEYRIT RRRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPDMGPDDF QTALDTRLGE LGTLRTRSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*

```

ORF19ng-1 (SEQ ID NO: 112) and ORF19-1 (SEQ ID NO: 106) show 95.5% identity in 716 aa overlap:

45  
50

```

              10      20      30      40      50      60
orf19-1.pep  MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
              |||||
orf19ng-1    MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf19-1.pep  NIITVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
              |||:|||||
orf19ng-1    NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
              70      80      90      100     110     120

```

5	orf19-1.pep	130	140	150	160	170	180
	orf19ng-1	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA					
10	orf19-1.pep	190	200	210	220	230	240
	orf19ng-1	DFFDPDAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ					
15	orf19-1.pep	250	260	270	280	290	300
	orf19ng-1	DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGGQACRNTAQALRASKDYVYSKRLG					
20	orf19-1.pep	310	320	330	340	350	360
	orf19ng-1	RAIEGCRQSLRLSDSNDSPDIRHLRRLDNLGSDVQQFRQLQHNGLQAENDRMGDTRIA					
25	orf19-1.pep	370	380	390	400	410	420
	orf19ng-1	ALETSSLKNTWQAIROPQNLNLESGVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFV					
30	orf19-1.pep	430	440	450	460	470	480
	orf19ng-1	CQPNYTATKSRVQRRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF					
35	orf19-1.pep	490	500	510	520	530	540
	orf19ng-1	STFFITIQAQTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL					
40	orf19-1.pep	550	560	570	580	590	600
	orf19ng-1	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ					
45	orf19-1.pep	610	620	630	640	650	660
	orf19ng-1	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF					
50	orf19-1.pep	670	680	690	700	710	
	orf19ng-1	QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHROPQNAAX					

In addition, ORF19ng-1 (SEQ ID NO: 112) shows significant homology to a hypothetical gonococcal protein (SEQ ID NO: 1121) previously entered in the databases:

```

5      sp|O333369|YOR2_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl|PID|e1154438
      (AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417
      Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203
      Identities = 301/326 (92%), Positives = 306/326 (93%)

      Query:   307 RQSLRLLSDGNDSPDIRHLSRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS 366
                RQSLRLLSDGNDSDIRHLSRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS
      Sbjct:    1  RQSLRLLSDGNDSDXDIRHLSRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS 60

10      Query:   367 FKNTWQAIROPQLNLESCVFRHAVRLSLVAAACTIVEALNNLGYWILLTALFVCQPNYT 426
                FKNTWQAIROPQLNLESVFRHAVRLSLVAAACTIVEALNNLGYWILLT LFVCQPNYT
      Sbjct:    61 FKNTWQAIROPQLNLESGVFRHAVRLSLVAAACTIVEALNNLGYWILLTRLFVCQPNYT 120

      Query:   427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAAGTTLFFMTRTRYKYSFSTFFIT 486
                ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAAGTTLFFMTRTRYKYSFSTFFIT
15      Sbjct:   121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAAGTTLFFMTRTRYKYSFSTFFIT 180

      Query:   487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTALERTAAALAVCSSG 546
                IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTALERTAAALAVCSSG
      Sbjct:   181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTALERTAAALAVCSSG 240

      Query:   547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADSLQPGFTLL 606
                TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFAD+ P
20      Sbjct:   241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADTCNPALPCS 300

      Query:   607 KTG YALTGYISALGAYRSEMHEECSP 632
                K  ALTGYISALG  ++  +  +P
25      Sbjct:   301 KPATALTGYISALGHTAAKCTKNAAP 326

```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein (SEQ ID NO: 1120), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 113):

```

35      1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
      51  GCGCGTTTGTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCCGGCG
      101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACT GCCCAACCTG
      151 CTTGCCCGCG TGTTCGCGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
      201 TTTGGCGGAA TACAAGGAAA CGCGTCAAA AGAGGCGG.C GAAGCCTTTA
      251 TCCGCCATGT GCGGGGATG CTGTCGTTG TACTGTTAT CGTTACCGCG

```

5  
10  
15  
20

```

301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
351 TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTT GCTGCGGATT
401 ACGTTTCCTT ATATATTATT GATTTCCTG TCTTCATTTG TCGGCTCGGT
451 ACTCAATTCT TATCATAAGT TCGGCATTCC GCGGTTTACG CCAC.GTTTC
501 TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTCGAT
551 CCGCCCGTTA CCGCGCyGCG GTGGGCGGTC TTTGTGCGCG GCATTTTGCA
601 ACTCGrmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
651 CCAAActGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
701 GCGCCTGCGA TTTTgGGCGT GaGCGTGCGC CAGGTTTCTT TGGTGATCAA
751 CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
801 ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GCGCGCACTC
851 GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
901 GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGcATGctgc
951 TGACGCTGCC GCGGgcGGTC GGACTGGCGG TGTGTGCGTT cCCgCtGGTG
1001 GCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC
1051 GCAACACGCG CTGATTGCCT ATTCTTTCCG TTTAATCGGC TTAATCATGA
1101 TACGATTTTc GGCACCCGCG TTCTATGCGC GGCAAAACAT CAaWAmGCCC
1151 GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
1201 CTTTAYCGGC CCACTrrAAC rCaTCGGAC TTTGCTTGC CATCGGTCTG
1251 GCGCGGTGTA TCAATGCCCG ATTGTTGTTT TACCTGTTGC GCAGACACGG
1301 TATTTACCAA CCTGG.CAAG GGTGGGCAG CGTTCCTT.AG CAAAAATGCT
1351 GcTCTCGCTC GCCGTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 114; ORF20):

25  
30  
35

```

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPSFQAD ADKFQLSIDL LRITPPYILL ISLSSFVGSV
151 LNSYHKFGIP AFTPXFLNVS FIVFALFFVP YFDPPTAXA WAVFVGILQ
201 LXFLPLWAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
401 FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
451 SRSP*

```

These sequences were elaborated, and the complete DNA sequence (SEQ ID NO: 115) is:

40  
45  
50  
55

```

1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
101 CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAAC GCCCAACCTG
151 CTTGCGCGCG TGTTTGCGGA GGGGGCGTTT GCCAAGCGT TTGTGCCGAT
201 TTTGGCGGAA TACAAGGAAA CGCGTTCAA AGAGGCGGCG GAGGCTTTTA
251 TCCGCCATGT GCGGGGATG CTGTCGTTG TACTGGTTAT CGTTACCGCG
301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
351 TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTG CTGCGGATTA
401 CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCAATTGT CCGCTCGGTA
451 CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTTCGTGCC TATTTCGATC
551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTCGCGCG CATTTTGCAA
601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
801 CGCCGACCGC ATGATGGAGC TGCCAGCGG CGTGCTGGGG GCGGCACTCG
851 GTACGATTTT CCGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG
901 GAACAGTTTT GCTGCTGCT CGACTGGGT TTGCGCCTGT GCATGCTGCT
951 GACGCTGCCG GCGGCGGTG GACTGGCGGT GTTGTGCTTC CCGCTGGTGG
1001 CGACGCTGTT TATGTACCGC GAATTTACGC TGTTTGACGC GCAGATGACC

```



5  
10  
15  
20  
25  
30  
35  
40  
45  
50

```

1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
1101 TAAAGTGTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCC
1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251 CGCGTGTATC AATGCCGGAT TGTTGTTTTC CCTGTTGCGC AGACACGGTA
1301 TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTAGCAAA AATGCTGCTC
1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC
1401 GTTTGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GCGCGCTTTG
1501 GGCTTCCGTC CGCGCCATT CAAACGCGTG GAAACTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 116; ORF20-1):

15  
20  
25  
30  
35  
40  
45  
50

```

1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPTALA WAVFVGILQ
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
501 GFRPRHFKRV EN*

```

25 Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169) (SEQ ID NO: 1122)

ORF20 (SEQ ID NO: 114) and MviN proteins (SEQ ID NO: 1122) show 63% aa identity in 440aa overlap:

30  
35  
40

```

Orf20 1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL LRRVFAEGAF 60
Mn+L +LA V S+TM SRVLGF RD ++AR FGAGMATDA FFVAFKLPNL LRR+FAEGAF
MviN 14 MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDA FFVAFKLPNL LRRIFAEGAF 73

Orf20 61 AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD 120
+QAFVPILAEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA
MviN 74 SQAFVPILAEYKSKQGEEATRIFVAYVSGLLTLALAVTVAGMLAAPWVIMVTAPGFADT 133

Orf20 121 ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP 180
ADKF L+ LLRITFPYILLISL+S VG++LN++++F IPAF P FLN+S I FALF P
MviN 134 ADKFALTTLRITFPYILLISLASLVGAILNTWNRFSIPAFAPFTFLNISMIGFALFAAP 193

Orf20 181 YFDPPTAXAWAVFVGILQLXFLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV 240
YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQM PAILGV
MviN 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRIINFRDTGAMRVVKQMGPAAILGV 253

Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT 300
SV+Q+SL+INTIFAS+L SGSVSWMYADR+ME PSGVLG ALGTILLP+LSK A+ +
MviN 254 SVSISLIINTIFASFLASGSVSWMYADRLMEFPPSGVLGVALGTILLPSLSKSFASGNH 313

```

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360  
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G  
 MviN 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYKGKFTAFDAAMTQRALIAYSVG 373

Orf20 361 LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXXCI 420  
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+  
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

Orf20 421 NAGLLFYLLRRHGIYQPXQG 440  
 NA LL++ LR+ I+ P G  
 MviN 434 NASLLYWQLRKQNIPTPQPG 453

# 10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 (SEQ ID NO: 114) shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) (SEQ ID NO: 118) from strain A of *N. meningitidis*:

15	orf20.pep	10 20 30 40 50 60	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
	orf20a	10 20 30 40 50 60	MNMLGALVKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
20	orf20.pep	70 80 90 100 110 120	AQAFVPILA EYKETS KEA XEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPSFAQD
	orf20a	70 80 90 100 110 120	AQAFVPILA EYKETS KEA EAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAKD
25	orf20.pep	130 140 150 160 170 180	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP
	orf20a	130 140 150 160 170 180	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFTPFTFLNVSFIVFALFFVP
30	orf20.pep	190 200 210 220 230 240	YFDPPVTAXAWAVFVG GILQLX FQLPWLAKLGFLKLPKLSFKDAAVNRVMQM A PA I LGV
	orf20a	190 200 210 220 230 240	YFDPPVTALAWAVFVG GILQLG FQLPWLAKLGFLKLPKLSFKDAAVNRVMQM A PA I LGV
35	orf20.pep	250 260 270 280 290 300	SVAQVSLVINTIFASYLQSGSVSWMYYADRMELPSGVLGAALGTILLPTLSKHSANQDT
	orf20a	250 260 270 280 290 300	SVAQISLVINTIFASYLQSGSVSWMYYADRMELPGGVLGAALGTILLPTLSKHSANQDT
40	orf20.pep	310 320 330 340 350 360	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG
	orf20a	310 320 330 340 350 360	EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG
45	orf20.pep	370 380 390 400 410 420	LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI

5

orf20a	<u>LIGLIMIKVLAPGPFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI</u>
	370 380 390 400 410 420
	430 440 450
orf20.pep	<u>NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSPX</u>
	:  ::  :
orf20a	<u>NAGLLFYLLRRHGIYQP GKGWAAFLAKMLLSLAVMGGGLYAAQIWL PFDWAHAGGMQKAA</u>
	430 440 450 460 470 480

The complete length ORF20a nucleotide sequence (SEQ ID NO: 117) is:

10	1	ATGAATATGC	TGGGAGCTTT	GGTAAAAAGTC	GGCAGCCTGA	CGATGGTGTCT
	51	GCGCGTTTTG	GGATTTGTGC	GCGATACGGT	CATTGCGCGC	GCATTGCGCG
	101	CAGGCATGGC	GACGGATGCG	TTCTTTGTCTG	CGTTCAAACCT	GCCCAACCTG
	151	CTTCGCCGCG	TGTTTTGCGGA	GGGGGCGTTT	GCCCAAGCGT	TTGTGCCGAT
	201	TTTGGCGGAA	TATAAGGAAA	CGCGTTCTAA	AGAGGCGACG	GAGGCTTTTA
15	251	TCCGCCATGT	GGCGGGGATG	CTGTCTGTTT	TACTGGTCAT	CGTTACCGCG
	301	CTGGGCATAC	TTGCCGCGCC	TTGGGTGATT	TATGTTTTCCG	CACCCGGTTT
	351	TGCCAAAGAT	GCCGACAAAT	TTCAGCTCTC	TATCGATTTG	CTGCGGATTA
	401	CGTTTTCTTA	TATCTTATTG	ATTTCACTTT	CCTCTTTTGT	CGGCTCGGTA
	451	CTCAATTCTT	ATCATAAATT	CAGCATTCCT	GCGTTTACGC	CCACGTTTCT
20	501	GAACGTGTCG	TTTATCGTAT	TCGCGCTGTT	TTTCGTGCCG	TATTTTCGATC
	551	CTCCCGTTAC	CGCGCTGGCT	TGGGCGGTTT	TTGTCCGCGG	CATTTTGCAA
	601	CTCGGCTTCC	AACTGCCCTG	GCTGGCGAAA	CTGGGTTTTT	TGAAACTGCC
	651	CAAAC TGAGT	TTCAAAGATG	CGGCGGTCAA	CCGCGTGATG	AAACAGATGG
	701	CGCCTGCGAT	TTTGGGCGTG	AGCGTGGCGC	AGATTTCTTT	GGTGATCAAC
25	751	ACGATTTTCG	CGTCTTATCT	GCAATCGGGC	AGCGTTTCAT	GGATGTATTA
	801	CGCCGACCGC	ATGATGGAAC	TGCCCGCGCG	CGTGCTGGGG	GCGGCATCTG
	851	TACAGATTTT	GCTGCCGACT	TTGTCCAAAC	ACTCGGCAAA	CCAAGATACG
	901	GAACAGTTTT	CCGCCCTGCT	CGACTGGGGT	TTGCCGNTGT	GCATGCTGCT
	951	GACGCTGCCG	GCGGCGGTG	GAATGGCGGT	GTGTGCTGTC	CCGCTGGTGG
30	1001	CAACCTTGTT	TATGTACCGA	GAATTCACGC	TGTTTGACGC	GCAGATGACG
	1051	CAACACGCGC	TGATTGCCTA	TTCTTTGCGT	TTAATCGGTT	TAATCATGAT
	1101	TAAAGTGTTG	GCGCCCGGCT	TTTATGCGCG	GCAAAACATC	AAAACGCCCC
	1151	TCAAAATCGC	CATCTTCACG	CTCATTTGCA	CGCAGTTGAT	GAACCTTGCC
	1201	TTTATCGGCC	CAC TGAACA	CGTCGGAATT	TCGCTTGCCA	TCGCTCTGGG
35	1251	CGCGTGATC	AATGCCGAT	TGTTGTTTTA	CCTGTTGCGC	AGACACGGTA
	1301	TTTACCAACC	TGGCAAGGGT	TGGGCAGCGT	TCTTGGCAAA	AATGCTGCTC
	1351	TCGCTCGCCG	TGATGGGAGG	CGGCCTGTAT	GCCGCCCAAA	TCTGGCTGCC
	1401	GTTTCGACTG	GCACACGCCG	GCGGAATGCA	AAAGGCCGCC	CGGCTCTTCA
	1451	TCCTGATTGC	CGTCGGCGGC	GGACTGTATT	TCGCATCACT	GGCGGCTTTG
40	1501	GGCTTCCGTC	CGCGCCATTT	CAAACGCGTG	GAAGCTGA	

This encodes a protein having amino acid sequence (SEQ ID NO: 118):

	1	MMNLGALVKV	GSLTMSRVRL	GFVRDTVIAR	AFGAGMATDA	FFVAFKLPNL
	51	LRRVFAEGAF	AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLVIVTA
45	101	<u>LGILAAPWVI</u>	YVSAPGFAKD	ADKFQLSIDL	LRITFPYILL	<u>ISLSSFVGSV</u>
	151	LNSYHKFSIP	AFTPTFLNVS	FIVFALFFVP	YFDPPTALA	WAVFVGILQ
	201	<u>LGFQLPWLAK</u>	LGFLKLPKLS	FKDAAVNRVM	<u>KQMAPAILGV</u>	<u>SVAQISLVIN</u>
	251	<u>TIFASYLQSG</u>	SVSWMYADR	MMELPGGVLG	AALGTILLPT	LSKHSANQDT
	301	EQFSALLDWG	LRXCMLLTLP	<u>AAVGMAVLSF</u>	PLVATLFMYR	EFTLFDAQMT
50	351	QHALIAYSFG	LIGLIMIKVL	APGFYARQNI	KTPVKIAIFT	LICTQLMNLA
	401	<u>FIGPLKHVGL</u>	<u>SLAIGLGACI</u>	<u>NAGLLFYLLR</u>	RHGIYQPGKG	<u>WAAFLAKMLL</u>
	451	<u>SLAVMGGGLY</u>	<u>AAQIWLFPDW</u>	<u>AHAGGMQKAA</u>	<u>RLFILIAVGG</u>	<u>GLYFASLAAL</u>
	501	GERPRHFKRV	ES*			

ORF20a (SEQ ID NO: 118) and ORF20-1 (SEQ ID NO: 116) show 96.5% identity in 512 aa overlap:

5	orf20a.pep	10 20 30 40 50 60	MNMLGALVKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
	orf20-1	10 20 30 40 50 60	MNMLGALAKVGS LTMVSRVLGFVRD TVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
10	orf20a.pep	70 80 90 100 110 120	AQAFVPILA EYKETRSKEATEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAKD
	orf20-1	70 80 90 100 110 120	AQAFVPILA EYKETRSKEAAEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPGFAQD
15	orf20a.pep	130 140 150 160 170 180	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAF TPTFLNVSFIVFALFFVP
	orf20-1	130 140 150 160 170 180	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAF TPTFLNVSFIVFALFFVP
20	orf20a.pep	190 200 210 220 230 240	YFDPPTALAWAVFVG GILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
	orf20-1	190 200 210 220 230 240	YFDPPTALAWAVFVG GILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
25	orf20a.pep	250 260 270 280 290 300	SVAQISLVINTIFASYLQSGSVSWMYADRM MELPGGVLGAALGTILLPTLSKHSANQDT
	orf20-1	250 260 270 280 290 300	SVAQVSLVINTIFASYLQSGSVSWMYADRM MELPSGVLGAALGTILLPTLSKHSANQDT
30	orf20a.pep	310 320 330 340 350 360	EQFSALLDWGLRXCM LTLPAAVGMAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG
	orf20-1	310 320 330 340 350 360	EQFSALLDWGLRLCML LTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG
35	orf20a.pep	370 380 390 400 410 420	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTL ICTQLMNLAFIGPLKHVGLSLAIGLGACI
	orf20-1	370 380 390 400 410 420	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTL ICTQLMNLAFIGPLKHVGLSLAIGLGACI
40	orf20a.pep	430 440 450 460 470 480	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWL PFDWAHAGGMQKAA
	orf20-1	430 440 450 460 470 480	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG
45	orf20a.pep	490 500 510	RLFILIAVGGGLYFASLAALGFRPRHFKRVESX
	orf20-1	490 500 510	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF20 (SEQ ID NO: 114) shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) (SEQ ID NO: 120) from *N. gonorrhoeae*:

5	orf20.pep	MNMLGALAKVGS LTMVSRVLGFVRD TVIARA FGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20ng	MNMLGALAKVGS LTMVSRVLGFVRD TVIARA FGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20.pep	AQAFVPILAEYK ETRSKEAXEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPSFAQD	120
	orf20ng	AQAFVPILAEYK ETRSKEATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD	120
10	orf20.pep	ADKFQLSIDLRLITFPYILLISLSS FVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
	orf20ng	ADKFQLSISLRLITFPYILLISLSS FVGSILNSYHKFGIPAFTPFTLNISFIVFALFFVP	180
15	orf20.pep	YFDPPTAXAWAVFVG GILQLXFPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV	240
	orf20ng	YFDPPTALAWAVFVG GILQLGFQPLWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV	240
	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMY YADRMELPSGVLGAALGTILLPTLSKHSANQDT	300
	orf20ng	SVAQISLVINTIFASYLQSGSVSWMY YADRMELPGGVLGAALGTILLPTLSKHSANQDT	300
20	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG	360
	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG	360
	orf20.pep	LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI	420
	orf20ng	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI	420
25	orf20.pep	NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP	454
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP	454

An ORF20ng nucleotide sequence (SEQ ID NO: 119) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 120):

35	1	MNMLGALAKV	GSLTMVSRVL	GFVRD TVIAR	AFGAGMATDA	FFVAFKLPNL
	51	LRRVFAEGAF	AQAFVPILAE	YKETRSKEAT	EAFIRHVAGM	LSFVLIVVTA
	101	LGILAAPWVI	YVSAPGFTKD	ADKFQLSISL	LRITFPYILL	ISLSSFVSGI
	151	LNSYHKFGIP	AFTPTFLNIS	FIVFALFFVP	YFDPPTALA	WAVFVGILQ
	201	LGFLQPLWAK	LGFLKLPKLN	FKDAAVNRVM	KQMAPAILGV	SVAQISLVIN
	251	TIFASYLQSG	SVSWMYADR	MMELPGGVLG	AALGTILLPT	LSKHSANQDT
	301	EQFSALLDWG	LRLCMLLTLP	AAAGLAVLSF	PLVATLFMYR	EFTLFDAQMT
	351	QHAIAYSFG	LIGLIMIKVL	ASGFYARQNI	KTPVKIAIFT	LICTQLMNL
40	401	FIGPLKHAGL	SLAIGLGACI	NAGLLFFLFR	KHGIYRPGQG	LGQPSWRKCC
	451	SRSP*				

Further DNA sequence analysis revealed the following DNA sequence (SEQ ID NO: 121):

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG  
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTGCGCG  
 101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACT GCCCAACCTG  
 151 CTTGCCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT  
 5 201 TTTGGCGGAA TATAAGGAAA CGCGTCTTAA AGAGGCGAcg gAGGCTTTTA  
 251 TCCGCCACGt tgcgggAatg CTGTCTTTG TGCTGATcgt cGttacCGCG  
 301 CTGGGCATAC TTGCCGCgcc tTGGGTGATT TATGTTtccg CgcccGGCTT  
 351 TACCAAAGAC GCGGACAAGT TCCAACTTTC CATCAGCCTG CTGCGGATTA  
 10 401 CGTTTCTTTA TATATTATG ATTTCTTTGT CTTCTTTGT CGGCTCGATA  
 451 CTCAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT  
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCTGTCCG TATTTTCGATC  
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTTT TTGTGCGCGG TATTTTCGAG  
 601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAGTGGC  
 651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG  
 15 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta  
 801 cgCCGACCGC ATGATGGAGc tgcgccGGGG CGTGCTGGGG GCTGCACTCG  
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG  
 901 GAACAGTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT  
 20 951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCTGTC CCGCTGGTGG  
 1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAAATGACG  
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATTATGAT  
 1101 TAAAGTGTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCCG  
 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC  
 25 1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCTGGG  
 1251 CGCGTGCATC AACGCCGGAT TGTTGTTCTT CCTGTTGCGC AAACACGGTA  
 1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC  
 1351 GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCCTGCC  
 1401 GTTCGAATGG GCGCACGCCG GCGGAATCGG GAAAGCGGG CAGCTCTGCA  
 30 1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCTCT GCGGCTTTG  
 1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

This encodes the following amino acid sequence (SEQ ID NO: 122; ORF20ng-1):

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL  
 35 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA  
 101 LGILAAPWVI YVSAPGF TKD ADKFQLSISL LRITFPYILL ISLSSFVGS  
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPTALA WAVFVGILQ  
 201 LGFQLPWLAK LGFLKLPKLN FKDAVNRVM KQMAPAILGV SVAQISLVIN  
 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT  
 40 301 EQFSALLDWG LRLCMLLTL AAAGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA  
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPRG WAAFLAKMLL  
 451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL  
 501 GFRPRHFKRV ES\*

ORF20ng-1 (SEQ ID NO: 122) and ORF20-1 (SEQ ID NO: 116) show 95.7% identity in 512 aa overlap:

	10	20	30	40	50	60
orf20-1.pep	MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
orf20ng-1	MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
	10	20	30	40	50	60
	70	80	90	100	110	120
orf20-1.pep	AQAFVPILAE	YKETRSKEA	EAFIRHVAG	MSFVLIVVT	ALGILAAPW	VIYVSAPG



Sbjct: 14 MNLLKSLAAVSSMTMFSRVLG FARD AIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73

Query: 61 AQAFVPILA EYK ETRSKEATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD 120  
+QAFVPILA EYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF

Sbjct: 74 SQAFVPILA EYKSKQGE EATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133

5 Query: 121 ADKFQLSISLLRITFPYILLISLSS FVGSILNSYHKFGIPAF TPTFLNISFIVFALFFVP 180  
ADKF L+ LLRITFPYILLISL+S VG+ILN++++F IPAF PTFNIS I FALF P

Sbjct: 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

Query: 181 YFDP PVTALAWAVFVG GILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV 240  
YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAILGV

10 Sbjct: 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLP RINFRDTGAMRVVKQMGPA I LGV 253

Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRM MELRRGV LGAALGTILLPTLSKHSANQDT 300  
SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ +

Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFP SGVLGVALGTILLPSLSKSFASGNH 313

Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 360  
+++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G

15 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQY GKFTA FDAAMTQRALIAYS VG 373

Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420  
LIGLI++KVLG GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+

Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

20 Query: 421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467  
NA LL++ LRK I+ P GW VM L+ +P

Sbjct: 434 NASLLYWQLRKQNI FT PQPGWMWFLMRIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220  
Identities = 14/41 (34%), Positives = 23/41 (56%)

25 Query: 469 EWAHAGGMRKAGQLCILI AVGGGLYFASLAALGFRPRHFKR 509  
EW+ + + +L ++ G YFA+LA LGF+ + F R

Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

Based on this analysis, including the homology with a virulence factor (SEQ ID NO: 1122) from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 15

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 123):

1 atGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
51 GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGCGC  
35 101 AAGAATATGC CGGTATGCGC CCTCGATGA AAGTCAAGGA AGGCGATGCC  
151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC C GTGGCGAAA  
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGCAA CGACGAAATC  
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAAC TTAA GCGGCGAAGA  
40 351 AGTGC GCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC



401 GTCCGTTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
451 GTCAATGCGA tGGACACCAA TCCG..

This corresponds to the amino acid sequence (SEQ ID NO: 124; ORF22):

5           1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA  
          51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEYNDEI  
         101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
         151 VNAMDTNP..

10 Further work revealed the complete nucleotide sequence (SEQ ID NO: 125):

          1 ATGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
         51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
         101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
         151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGCT  
15       201 GTTTACTGCG CCGGCTTCAG GCAAATCGC CGCGATTAC CGTGGCGAAA  
         251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
         301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTTA GCGGCGAAGA  
         351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC  
         401 GTCCGTTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
20       451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATTAT  
         501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT  
         551 TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG  
         601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC  
         651 TGCCGGTTTG AGTGGCACGC ACATTCATTT CATCGAGCCG GTCGGCGCGA  
25       701 ATAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT  
         751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG  
         801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTGG GGTGCGAAAG  
         851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT  
         901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT  
30       951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
         1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAATACTC CATCACGCGT  
         1051 ACAACCTCG GCCATTCCTT GAAAAACAAA CTCTTCAAGT TCAACACAGC  
         1101 CGTCAACGGC GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG  
         1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC  
35       1201 GGCATACCG ACAGCGCGCA GGCATTGGGT TGCTTGGAAT TGGACGAAGA  
         1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCG GGGCAAATAC GAATACGGCC  
         1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence (SEQ ID NO: 126; ORF22-1):

40           1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA  
          51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI  
         101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
         151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV  
45       201 SENAANIETH EFGGPHAGL SGTHIHFI EP VGANKTVWTI NYQDVITIGR  
         251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI  
         301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPKYSITR  
         351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLRLRLDILV  
         401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG\*

50 Further work identified the corresponding gene in strain A of *N.meningitidis* (SEQ ID NO: 127):

          1 ATGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
         51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG

5  
10  
15  
20  
25

```

101 AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAAGNATC CGGGCGTGTT
201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA
251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACTTAA GCGGCGANGA
351 ANTNNNGNNGC AATCTGATCC AATCCGTTT GTGGACTGCG CTGCGTANCC
401 GTCCGTTTCA CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT
501 CAAAGAAGCC GNCGANGAT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
551 TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
651 GGCCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA
701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
801 TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTGG GGTGCGAAAG
851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
901 TCCGGTTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAAATCTC CATCACGCGT
1051 ACGACCCTCG GCCATTTCTT GAAAAACAAA CTCTTCAAGT TCACGACAGC
1101 CGTCAACGGT GGCAGCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
1201 GGCATACCG ACAGCGCGCA AGCATTGGGT TGCTTGAAT TGGACGAAGA
1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 128; ORF22a):

30  
35

```

1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMKVKEGDA
51 VKKGQVLFED KXXPGVVFTA PVSGKIAAIH RGEKRVLQSV VIAVEGNDEI
101 EFERYAPEAL ANLSGXEXX NLIQSLWTA LRXPFSKIP AVDAEPFAIF
151 VNAMDTNPLA ADPVVVIKEA XXDFRXXLV LSRLTERKIH VCKAAGADVP
201 SENAANIETH EFGGPHAGL SGTHIHFI EP VGANKTVWTI NYQDVIAIGR
251 LFATGRLNTE RVIALGGSQV NKPRLLRVL GAKVSQITAG ELVDADNRVI
301 SGSVLNGAIT QGAHDYLG RY HNQISVIEEG RSKELFGWVA PQPDKYSITR
351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLRLDLIV
401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

```

The originally-identified partial strain B sequence (ORF22) (SEQ ID NO: 124) shows 94.2% identity over a 158aa overlap with ORF22a (SEQ ID NO: 128):

40  
45  
50

```

          10      20      30      40      50      60
orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
          |||||:||||:|||||
orf22a    MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED
          10      20      30      40      50      60

          70      80      90      100     110     120
orf22.pep KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEXNDEIEFERYAPEALANLSGEEVRR
          |||||:|||||
orf22a    KXXPGVVFTAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGXEXX
          70      80      90      100     110     120

          130     140     150
orf22.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP
          |||||:|||||
orf22a    NLIQSLWTALRXPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV

```

130 140 150 160 170 180

The complete strain B sequence (ORF22-1) (SEQ ID NO: 126) and ORF22a (SEQ ID NO: 128) show 94.9% identity in 447 aa overlap:

```

5      10      20      30      40      50      60
orf22a.pep  MIKIKKGLNLP IAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED
orf22-1      MIKIKKGLNLP IAGRPEQAVYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
           10      20      30      40      50      60

10     70      80      90     100     110     120
orf22a.pep  KXXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX
orf22-1      KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR
           70      80      90     100     110     120

15     130     140     150     160     170     180
orf22a.pep  NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV
orf22-1      NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV
           130     140     150     160     170     180

20     190     200     210     220     230     240
orf22a.pep  LSR LTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI
orf22-1      LSR LTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI
           190     200     210     220     230     240

25     250     260     270     280     290     300
orf22a.pep  NYQDVIAIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADNRVI
orf22-1      NYQDVITIGRLFATGRLNTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDTDNRVI
           250     260     270     280     290     300

30     310     320     330     340     350     360
orf22a.pep  SGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK
orf22-1      SGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK
           310     320     330     340     350     360

35     370     380     390     400     410     420
orf22a.pep  LFKFTTAVNGGDRAMVPIGTYERVMLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA
orf22-1      LFKFNTAVNGGDRAMVPIGTYERVMLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA
           370     380     390     400     410     420

40     430     440
orf22a.pep  LCSFVCPGKYEXGPLL RKVLETXEKEGX
orf22-1      LCSFVCPGKYEYGPLL RKVLETIEKEGX
           430     440

```

45 Further work identified a partial gene sequence (SEQ ID NO: 129) from *N.gonorrhoeae*, which encodes the following amino acid sequence (SEQ ID NO: 130; ORF22ng):

5

1	MIKIKKGLNL	PIAGRPEQVI	YDGAITEVA	LLGEEYVGM	PSMKIKEGEA
51	VKKGQVLFED	KNPNVVF	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF
151	VNAMDTNPLA	ADPTVIIKEA	AEDFKRGLLV	LSRLTERKIH	VCKAAGADVP
201	SENAANIETH	EFGGPHAPGL	SGTHIHFI	VGANKTVWTI	NYQDVIAIGR
251	LFVTGRNLTE	RVALGGLQV	NKPRLLR	TVL	GAKVSLQTA
301	SGSVLNGAIA	OGAHDYLG	RY	HN*	ELVDADNRVI

Further work identified complete gonococcal gene (SEQ ID NO: 131):

10	1	ATGATTAAAA	TCAAAAAAAG	TCTAAATCTG	CCCATCGCGG	GCAGACCGGA
	51	GCAAGTCATT	TATGACGGCC	CGGCCATTAC	CGAAGTCGCG	TTGCTTGGCG
	101	AAGAATATGT	CGGCATGCGC	CCCTCGATGA	AAATCAAGGA	AGGTGAAGCC
	151	GTCAAAAAAG	GCCAAGTGCT	GTTTGAAGAC	AAAAAGAATC	CGGGCGTAGT
	201	ATTTACTGCG	CCGGCTTCAG	GCAAAATCGC	CGCTATTAC	CGTGGCGAAA
15	251	AGGCGGTACT	TCAGTCAGTC	GTGATTGCGC	TTGAAGGCAA	CGACGAAATC
	301	GAGTTCGAAC	GACTAGTACC	TGAAGCGCTG	GCAAAATTGA	GCAGCGAAAA
	351	AGTGCGCCGC	AACCTGATT	AATCAGGCTT	ATGGACTGCG	CTTCGCACCC
	401	GTCCGTTCAG	CAAAATCCCT	GCCGTAGATG	CCGAGCCGTT	CGCCATCTTC
	451	GTCAATGCGA	TGGACACCAA	TCCGCTGGCT	GCCGACCCTA	CGGTCAATCAT
20	501	CAAAGAAGCC	GCCGAAGACT	TCAAACGCGG	CCTGTTGGTA	TTGAGCCGCC
	551	TGACCGAAGC	TAAAATCCAT	GTGTGTAAAG	CAGCAGGCGC	AGACGTGCCG
	601	TCTGAAAAAT	CTGCCAATAT	CGAAACACAT	GAATTTGGCG	CGCCGCATCC
	651	TGCGCGCTTG	AGTGGCACGC	ACATTCATTT	CATTCGAGCA	GTCCGGCGCA
	701	ATAAAACCGT	GTGGACCATC	AATTATCAAG	ACGTGATTGC	TATCGGACGT
25	751	TTGTTCGTAA	CAGGCCGTCT	GAATACCGAG	CGCGTGTTG	CCTTGGGCGG
	801	CCTGCAAGTC	AACAAACCGC	GCCTCTTGCG	TACCGTTTTG	GGTGCGAAGG
	851	TGTCTCAACT	TACCGCCGGC	GAATTGGTTG	ACGCGGACAA	CCGCGTGATT
	901	TCCGGTTCCG	TATTGAACGG	TGCGATTGCA	CAGGCGCGC	ATGATTATTT
	951	GGGACGTAC	CACAATCAGA	TTTCCGTTAT	CGAAGAAGGC	CGCAGCAAAG
30	1001	AGCTGTTCCG	CTGGGTTGCG	CCGCAGCCGG	ACAAATACTC	CATCACGCGC
	1051	ACCACTCTCG	GCCATTTCTT	AAAAACAAA	CTCTTCAAGT	TCACGACAGC
	1101	CGTCAACGCG	GGCGACCGCG	CCATGTATCC	GATCGGCACT	TATGAGCGCG
	1151	TAATGCCGTT	GGACATCCTG	CCTACCTTGC	TTTTGCGCGA	TTTAATCGTC
	1201	GGCGATACCG	ACAGCGCGCA	GGCTTTGGGT	TGCTTGAAT	TGGACGAAGA
35	1251	AGACCTCGCT	TTGTGACGCT	TCGTCTGCC	GGGCAAAATC	GAATACGGCC
	1301	CGCTGTTGCG	CAAAGTGCTG	GAAACCATTG	AGAAGGAAGG	CTGA

This encodes a protein having amino acid sequence (SEQ ID NO: 132; ORF22ng-1):

40	1	MIKIKKGLNL	PIAGRPEQVI	YDGPAITEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF
	151	VNAMDTNPLA	ADPTV I KEA	AEDFKRGLLV	LSRLTERKIH	VCKAAGADVP
	201	SENAANIETH	EFGGPHAGL	SGTHIHFIEP	VGANKTVWTI	NYQDVIAIGR
45	251	LFVTGRNLTE	RVALGGLQV	NKPRLLRVT	GAKVSQLTG	ELVDADNRVI
	301	SGSVLNGAIA	QGAHDYLG	HNQISVIEG	RSKELFGWA	PQPDKYSITR
	351	TTLGHFLKNK	LFKFTTAVNG	GDRAMVPIGT	YERVMLPLDIL	PTLLLRDLIV
	401	GDTDSAQALG	CLELDEEDLA	LCSFVCPGKY	EYGPLLRKVL	ETIEKEG*

The originally-identified partial strain B sequence (ORF22) (SEQ ID NO: 124) shows 93.7% identity over a 158aa overlap with ORF22ng (SEQ ID NO: 130):

50	orf22.pep	MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFD	60
		:           :           :	
	orf22nq	MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFD	60

5	orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVE XNDEIEFERYAPEALANLSGEEVRR	120
	orf22ng	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVE GND EIEFERYVPEALAKLSSEKVR	120
	orf22.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP	158
	orf22ng	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV	180

The complete sequences from strain B (ORF22-1) (SEQ ID NO: 126) and gonococcus (ORF22ng-1) (SEQ ID NO: 132) show 96.2% identity in 447 aa overlap:

10	orf22-1.pep	MIKIKKGLNLP IAGRPEQAVYDGP AITEVALLGEEYAGMRPSMKVKEGDAVKKQVLFED	10 20 30 40 50 60
	orf22ng-1	MIKIKKGLNLP IAGRPEQVIYDGP AITEVALLGEEYVGM RPSMKI KEGEAVKKQVLFED	10 20 30 40 50 60
15	orf22-1.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVE GND EIEFERYAPEALANLSGEEVRR	70 80 90 100 110 120
	orf22ng-1	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVE GND EIEFERYVPEALAKLSSEKVR	70 80 90 100 110 120
20	orf22-1.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV	130 140 150 160 170 180
	orf22ng-1	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV	130 140 150 160 170 180
25	orf22-1.pep	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI	190 200 210 220 230 240
	orf22ng-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI	190 200 210 220 230 240
30	orf22-1.pep	NYQDVITIGRLFATGRLNTERVIALGGSQV NKPRLLR TVLGAKVSQITAGELVD TNRVI	250 260 270 280 290 300
	orf22ng-1	NYQDVIAIGRLFVTGRLNTERVV ALGGLQV NKPRLLR TVLGAKVSQLTAGELVDADNRVI	250 260 270 280 290 300
35	orf22-1.pep	SGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK	310 320 330 340 350 360
	orf22ng-1	SGSVLNGAIAQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK	310 320 330 340 350 360
40	orf22-1.pep	LFKFNTAVNGDRAMVPIGTYERV MPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA	370 380 390 400 410 420
	orf22ng-1	LFKFNTAVNGDRAMVPIGTYERV MPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA	370 380 390 400 410 420
45	orf22-1.pep	LCSFVCPGKYEYG PLLRKVLETIEKEGX	430 440
	orf22ng-1	LCSFVCPGKYEYG PLLRKVLETIEKEGX	

430

440

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of Actinobacillus pleuropneumoniae (accession number U24492) (SEQ ID NO: 1123).

ORF22 (SEQ ID NO: 124) and this 48kDa protein (SEQ ID NO: 1123) show 72% aa identity in 158aa overlap:

```

Orf22  1  MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60
        MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFED
48kDa  1  MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

orf22  61  KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR 120
        KKNPGVVFTAPASG + I+RGEKRVLSVVI VE +++I F RY LA+LS E+V++
48kDa  61  KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

orf22  121 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
        NLI+SGLWTA RTRPFSPK+PA+DA P +IFVNAMDTNP
48kDa  121 NLIESGLWTAFRTRPFSPKVPALDAIPSSIFVNAMDTNP 158

```

ORF22a (SEQ ID NO: 128) also shows homology to the 48kDa *Actinobacillus pleuropneumoniae* protein (SEQ ID NO: 1123):

```

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]
Length = 449

```

```

Score = 530 bits (1351), Expect = e-150
Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

```

```

Query: 1  MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60
        MI IKKGL+LPIAG P QVI++G + EVA+LGEEY GMRP MKV+EGD VKKGQVLFED
Sbjct: 1  MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

```

```

Query: 61  KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX 120
        KK PGVVFTAP SG + I+RGEKRVLSVVI VEG+++I F RY LA+LS +
Sbjct: 61  KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

```

```

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNP LAADPVVVIKEAXXDFRXXLV 180
        NLI+SGLWTA R RPFSK+PA+DA P +IFVNAMDTNP LAADP VV+KE DF+ V
Sbjct: 121 NLIESGLWTAFRTRPFSPKVPALDAIPSSIFVNAMDTNP LAADPEVVLKEYETDFKDGTLTV 180

```

```

Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 237
        L+RL ++ +++CK A +++P S I F G HPAGL GTHIH++PVGA K V
Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

```

```

Query: 238 WTINYQDVIAIGRLFATGRINTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297
        W +NYQDVIAIG+LF TG L T+R+I+L G QV PRL+RT LGA +SQ+TA EL +N
Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRLGANLSQLTANELNAGEN 300

```

Query: 298 RVISGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 357  
 RVISGSVL+GA G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF  
 Sbjct: 301 RVISGSVLSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

5

Query: 358 KNKLFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXVGDTSQAQXXXXXXXXXX 417  
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ  
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTSQAQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGPLLRKVLETXEKEG 447  
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 (SEQ ID NO: 132) also shows homology with the OMP (SEQ ID NO: 1123) from

10 *A.pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus  
 pleuropneumoniae] Length = 449  
 Score = 555 bits (1414), Expect = e-157  
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

15

Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFED 86  
 MI IKKGL+LPIAG P QVI++G + EVA+LGEYVGMRPSMK++EG+ VKKGQVLFED  
 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

20

Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146  
 KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++  
 Sbjct: 61 KKNPGVVFTAPASGTVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 147 NLIQSGLWTA LRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAEDFKRGLLV 206  
 NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V++KE DFK GL V  
 Sbjct: 121 NLIESGLWTA FRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

25

Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 263  
 L+RL ++ +++CK A +++P S I F G HPAGL GTHIH++PVGA K V  
 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRNLNTERVVALGGLQVKNPRLRLTVLGAQVSQLTAGELVDADN 323  
 W +NYQDVIAIG+LF TG L T+R+++L G QV PRL+RT LGA +SQLTA EL +N  
 Sbjct: 241 WHLNQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRLGANLSQLTANELNAGEN 300

30

Query: 324 RVISGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFL 383  
 RVISGSVL+GA A G DYLGRY Q+SV+ EGR KELFGW+ P DK+SITRT LGHF  
 Sbjct: 301 RVISGSVLSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

35

Query: 384 KNKLFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXVGDTSQAQXXXXXXXXXX 443  
 K KLF FTTAV+GG+RAMVPIG YERVM GDTDSAQ  
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLLRDLAAGDTSQAQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473  
 ++VCPGK YGP+LR LE IEKEG  
 Sbjct: 420 DLALCTYVCPGKNNGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein (SEQ ID NO:

40 1123) of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis*

and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (SEQ ID NO: 126) (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 (SEQ ID NO: 126) is a surface-exposed protein, and that it is a useful immunogen.

## 10 Example 16

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 133):

```

1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAAC TG
51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
101 TTA CTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
15  151 TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20  401 TTTTGC GCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGS GC TTTmTTTGsw CAKcATCTTT TTTGCCGCAC AGTTTGTGCG
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCCTTAA AGAAGTCGGC TTGGGCGGCA GCGTGTTGTT TATCGGTTTT
601 ATTTTAATTT GTCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
25  651 ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGGCGCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
801 GrkCmmmmTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
30  901 TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951 TCCCGCACCT TAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 134; ORF12):

```

1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
35  51 SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201 ILICAFINLM IGSASAQWAV TAPIFVPMML LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
40  301 WV FVLGLPVG PGAPTFFYPAP *

```

Further sequence analysis revealed the complete DNA sequence (SEQ ID NO: 135) to be:



1 ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA  
 51 ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTT ATTATTTTCA  
 101 TTGTGTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC  
 151 GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT  
 5 201 GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTATC AAAATCCTGA  
 251 CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCGTTGGG AACGGTGTG  
 301 GTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC  
 351 ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAATC ACTACTTTTA  
 401 TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT  
 10 451 GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGCCGCCA  
 501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT  
 551 CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC  
 601 CAACAGCGCG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC  
 651 CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT  
 15 701 ATTTTGTAC TGA AAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA  
 751 GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC  
 801 TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGGT GTTGCCTTAT  
 851 CCGCCCTATT GGCTTGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT  
 901 CCTGAAACAG GATTGGTTTC CGGTCGCCG TTTTAAAT CGATTGTTGT  
 20 951 TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATGTTTAT GGCCGGGTAA  
 1001 CCCGAAGTTT GCGCGCGGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG  
 1051 ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTGT CCGCACAGTT  
 1101 TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG  
 1151 GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTATTATC  
 25 1201 GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC  
 1251 CGCGCAATGG GCGGTAACTG CGCCGATTTT CGTCCCTATG CTGATGTTGG  
 1301 CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC  
 1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTTCCGGC TGATTATGGC  
 1401 GACGGTGATC AAATACAAAA AAGATGCGGG CGTGGGTACG CTGATTTCTA  
 30 1451 TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC  
 1501 TGCATTGGG TATTTGTTTT GGGCCTGCCG GTCGGTCCCG GCGCGCCAC  
 1551 ATTCTATCCC GCACCTTAA

This corresponds to the amino acid sequence (SEQ ID NO: 136; ORF12-1):

35 1 MSQTDTRQD RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS  
 51 VPDPRPVGAK GRADDGLIYI VSLLNADGFI KILHTVKNF TGFAPLGTVL  
 101 VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY  
 151 VVLIPLSAII FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT  
 201 QQAAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS  
 40 251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH  
 301 PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES  
 351 MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLEFI  
 401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS  
 451 VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF  
 45 501 CIWVFVLGLP VGPGAPTFYP AP\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 (SEQ ID NO: 134) shows 96.3% identity over a 320aa overlap with an ORF (ORF12a)

50 (SEQ ID NO: 138) from strain A of *N. meningitidis*:

5	orf12.pep	AXXI IHPXXVVGPEANWFFMVASTFVIALI
	orf12a	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALI
10	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
	orf12a	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
15	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
	orf12a	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
20	orf12.pep	TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGILICAFINLM
	orf12a	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGILICAFINLM
25	orf12.pep	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY
	orf12a	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY
30	orf12.pep	KKDAGVGTLIXMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAPX
	orf12a	KKDAGVGTLISMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAPX

The complete length ORF12a nucleotide sequence (SEQ ID NO: 137) is:

35	1	ATGAGTCAAA	CCGATACGCA	ACGGGACGGA	CGATTTTAC	GCACAGTCGA
	51	ATGGCTGGGC	AATATGTTGC	CGCACCCGGT	TACGCTTTT	ATTATTTTCA
40	101	TTGTGTTATT	GCTGATTGCC	TCTGCCCGCG	GTGCGTATTT	CGGACTATCC
	151	GTCCCCGATC	CGCGCCCTGT	TGGTGCGAAA	GGACGTGCCG	ATGACGGTTT
45	201	GATTCACGTT	GTCAGCCTGC	TCGATGCTGA	CGGTTTGATC	AAAATCCTGA
	251	CGCATACCGT	TAAAAATTC	ACCGGTTTCG	CGCCGTGGG	AACGGTGTTG
50	301	GTTTCTTTAT	TGGGCGTGGG	GATTGCGGAA	AAATCGGGCT	TGATTTCCGC
	351	ATTAATGCGC	TTATTGCTCA	CAAAATCTCC	ACGCAAATC	ACTACTTTTA
	401	TGGTTGTTTT	TACAGGGATT	TTATCTAATA	CCGCTTCTGA	ATTGGGCTAT
	451	GTCGCTCTAA	TCCCTTGTG	CGCCATCATC	TTTCATTCCC	TCCGCCGCCA
	501	TCCGCTTGCC	GGTCTGCTG	CGGCTTTCGC	CGGCGTTTCG	GGCGGTTATT
	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTTGGC	AGGCATCACC
	601	CAACAGCGCG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC
	651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
	701	ATTTTGTAC	TGAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
	751	GATTTGTAC	AAGAAGAAA	AGACATTCGA	CATTCCAATG	AAATCACGCC
	801	TTTGGAATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTT	GTTGCCTTAT
	851	CCGCCCTATT	GGCTTGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCCTCAT
	901	CCTGAAACAG	GATTGGTTTC	CGGTTCCGCG	TTTTTAAAT	CAATTGTTGT
	951	TTTTATTTTC	TTGTTGTTTG	CACTGCCGGG	CATTGTTTAT	GGCCGGGTAA
	1001	CCCGAAGTTT	GCGCGGCGAA	CAGGAAGTCG	TTAATGCGAT	GGCCGAATCG

5	1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTTTG	CCGCACAGTT
	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTTAAAG
	1151	GGGCGACGTT	CTTAAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCCCTC
	1251	CGCGCAATGG	GCGGTAAC TG	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
	1301	CCGGCTACGC	GCCCCAAGTC	ATTCAGCCG	CTTACCGCAT	CGGTGATTCC
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCCGGC	TGATTATGGC
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTTCTA
10	1451	TGATGTTGCC	GTATTCCGCT	TTCTTCTTGA	TGCGTGGAAT	TGCCTTATTC
	1501	TGCATTTGGG	TATTTGTTTT	GGGCTGCC	GTCGGTCCCG	GCGCGCCAC
	1551	ATTCTATCCC	GCACCTTAA			

This encodes a protein having amino acid sequence (SEQ ID NO: 138):

15	1	MSQTD	TQRDG	RFLRT	VEWLG	NMLPH	PVTLF	IIFIV	LLLLIA	SAAGAY	FGLS
	51	VPDPR	VGAK	GRADD	GLIHV	VSLLD	ADGLI	KILTHT	TVKNF	TGFAP	LGTVL
	101	VSLLG	VGIAE	KSGLI	SALMR	LLLT	KSPRKL	TTFMV	VFTGI	LSNTA	SELGY
	151	VVLIPL	SAII	FHSLG	GRHPLA	GLAA	FAGVS	GGYAN	LNFLG	TIDPL	LAGIT
	201	QQAQI	IHPD	YVVG	PEANWF	FMV	ASTFVIA	LIGYF	VTEKI	VEPQL	GPYQS
20	251	DLSQ	EKDIR	HSNEI	TPLEY	KGLIW	AGVVF	VALS	ALLAWS	IVPAD	GILRH
	301	PETGL	VSGSP	FLKSI	VVFIF	LLFAL	PGIVY	GRVTR	SLRGE	QEVVN	MAAES
	351	MSTL	GLYLV	IFFAA	QVFAV	FNW	TNIGQYI	AVKG	ATFLKE	VGLGG	SVLFI
	401	GFILIC	AFIN	LMIGS	SASAQW	AVT	APIFVPM	LMLAG	YAPEV	IQAA	YRIGDS
	451	VTNII	TMMS	YFGLI	MA TVI	KYK	KDAGVGT	LIS	MMLP	SA	FFLIAWIALF
	501	CIWVF	VFLGLP	VGPG	APT	FYP	AP*				

25 ORF12a (SEQ ID NO: 138) and ORF12-1 (SEQ ID NO: 136) show 99.0% identity in 522 aa overlap:

		10	20	30	40	50	60
	orf12a.pep	MSQTD	QRDGRFLRTVEWLG	NMLPHPVT	LFII FIV	LLLIASAAGAY	FGLSVPDPRPVGAK
30	orf12-1	MSQTD	QRDGRFLRTVEWLG	NMLPHPVT	LFII FIV	LLLIASAVGAY	FGLSVPDPRPVGAK
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf12a.pep	GRADDGLIHV	SLLDADGLIK	ILTHTVKN	FTGFAPLGT	VLVLSLLGV	GIAEKSGGLISALMR
35	orf12-1	GRADDGLIY	IVSLLNADG	FIKILTHTV	KNFTGFAPL	GTVLVLSLL	GVGIAEKSGGLISALMR
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf12a.pep	LLLT	KSPRKL	TTFMV	FTGILS	NTASEL	GYVVLIPLSAII FHS
40	orf12-1	LLLT	KSPRKL	TTFMV	FTGILS	NTASEL	GYVVLIPLSAII FHS
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf12a.pep	GGYSANL	FLGTID	PLL	AGITQQA	AI IHPDY	VVGPEANWFFMVASTFVIALIGYFVTEKI
45	orf12-1	GGYSANL	FLGTID	PLL	AGITQQA	AI IHPDY	VVGPEANWFFMVASTFVIALIGYFVTEKI
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf12a.pep	VEPQLG	PYQSDLS	QEEKDIR	HSNEITP	LEYKGLI	WAGVVFVLSALLAWSI

5	orf12-1	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH	250	260	270	280	290	300
	orf12a.pep	PETGLVSGSPFLKSIVVFI FLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI	310	320	330	340	350	360
10	orf12-1	PETGLVSGSPFLKSIVVFI FLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI	310	320	330	340	350	360
	orf12a.pep	IFFAAQFVAFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW	370	380	390	400	410	420
15	orf12-1	IFFAAQFVAFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW	370	380	390	400	410	420
	orf12a.pep	AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKGAGVGT	430	440	450	460	470	480
20	orf12-1	AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKGAGVGT	430	440	450	460	470	480
	orf12a.pep	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAPX	490	500	510	520		
	orf12-1	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAPX	490	500	510	520		

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 (SEQ ID NO: 134) shows 92.5% identity over a 320aa overlap with a predicted ORF

(ORF12.ng) (SEQ ID NO: 140) from *N. gonorrhoeae*:

30	orf12.pep	AXXI IHPXXVVGPEANWFFMVASTFVIALI	30
	orf12ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMAASTFVIALI	232
35	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	90
	orf12ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	292
40	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFI FLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150
	orf12ng	PADGILRHPETGLVAGSPFLKSIVVFI FLLFALPGIVYGRITRSLRGEREVVNAMAESMS	352
	orf12.pep	TLXLXLXXIFFAAQFVAFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM	210
	orf12ng	TLGLYLVI IFFAAQFVAFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIFIGFILICAFINLM	412
	orf12.pep	IGSASAQWAVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY	270
	orf12ng	IGSASAQWAVTAPIFVPMLMLAGNAPQVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY	472
	orf12.pep	KKDAGVGT LIXMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAP	320
	orf12ng	KKDAGVGT LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYPVP	522

The complete length ORF12ng nucleotide sequence (SEQ ID NO: 139) is:

```

1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
51 ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
101 TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATTT CGGACTATCC
5  151 GTCCCCGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
201 GATTCACGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
251 CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTG
301 GTTTCCTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
10  351 ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAATC ACTACTTTTA
401 TGGTTGTTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
451 GTCGTCCTAA TCCCTTTGTC CGCCGTCATC TTTCAATCGC TCGGCCGCCA
501 TCCGCTTGCC GGTTCGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
551 CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTGGC AGGCATCACC
601 CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
15  651 CAACTGGTTC TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
701 ATTTTGTAC TGAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
751 GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCAGCC
801 TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGTTT GTTGCCTTAT
851 CCGCCCTATT GGCTTGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
20  901 CCTGAAACAG GATTGGTTGC CGGTCGCCG TTTTAAAAAT CGATTGTTGT
951 TTTTATTTTC TTGTTGTTTG CGTGCCGGG CATTGTTTAT GGCCGATAA
1001 CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
1051 ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
1101 TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTAAAG
25  1151 GGGCGGTGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATTATC
1201 GGTTCCTATT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
1251 CGCGCAATGG GCGGTAATG CGCCGATTTT CGTCCCTATG CTGATGTTGG
1301 CCGGCTACGC GCCGGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTCGGGC TGATTATGGC
30  1401 GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCACG CTGATTCTA
1451 TGATGTTGCC GTATTCCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
1501 TGCATTTGGG TATTTGTTT GGGTCTGCCC GTCGGTCCC GCACACCCAC
1551 ATTCTATCCG GTGCCTTAA

```

35 This encodes a protein having amino acid sequence (SEQ ID NO: 140):

```

1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
51 VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
101 VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
40  151 VVLIPLSAVI FHS LGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
201 QQAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301 PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
351 MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAFLKK FRLGGSVLFI
401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
45  451 VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
501 CIWVFVLGLP VPGTPTFPY VP*

```

ORF12ng (SEQ ID NO: 140) shows 97.1% identity in 522 aa overlap with ORF12-1 (SEQ ID NO: 136):

```

50  orf12-1.pep      10      20      30      40      50      60
      MSQTDQTDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
      ||||:|:|||||
orf12ng MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK

```

-168-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf12-1.pep	GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSLGVGIAEKSGLISALMR					
	orf12ng	GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLGVGIAEKSGLISALMR					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf12-1.pep	LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS					
	orf12ng	LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAVIFHSLSGRHPLAGLAAAFAGVS					
		130	140	150	160	170	180
		190	200	210	220	230	240
15	orf12-1.pep	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
	orf12ng	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI					
		190	200	210	220	230	240
		250	260	270	280	290	300
20	orf12-1.pep	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
	orf12ng	VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
		250	260	270	280	290	300
		310	320	330	340	350	360
25	orf12-1.pep	PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNMAESMSTLGLYLVI					
	orf12ng	PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNMAESMSTLGLYLVI					
		310	320	330	340	350	360
		370	380	390	400	410	420
30	orf12-1.pep	IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFGFILICAFINLMIGSASAQW					
	orf12ng	IFFAAQFVAFFNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFGFILICAFINLMIGSASAQW					
		370	380	390	400	410	420
		430	440	450	460	470	480
35	orf12-1.pep	AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKGAVGT					
	orf12ng	AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKGAVGT					
		430	440	450	460	470	480
		490	500	510	520		
40	orf12-1.pep	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYPAPX					
	orf12ng	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYPVPX					
		490	500	510	520		

In addition, ORF12ng (SEQ ID NO: 140) shows significant homology with a hypothetical protein (SEQ ID NO: 1124) from *E.coli*:

45 sp|P46133|YDAH\_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION  
 )gi|1787597 (AE000231) hypothetical protein in ogt 5' region [Escherichia coli]  
 Length = 510  
 Score = 329 bits (835), Expect = 2e-89

Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)

Query: 8 RSGRFLRTVEWLGNMPLHPVTTTTTTTTTASAVGAYFGLSVDPDRPVGAKGRADDGL 67  
 +SG+ VE +GN +PHP +A+ + FG+S +P D  
 Sbjct: 13 QSGKLYGWVERIGNKVPHPFLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64

5 Query: 68 IHVVSLLDADGLIKILHTVKNFTGFAPXXXXXXXXXXXXXIAEKSGLISALMRLLLTKSP 127  
 + V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + +  
 Sbjct: 65 VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLPALMVKMASHVN 124

10 Query: 128 RKLTFMVVFTGILSNTASELGYVVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187  
 + ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL  
 Sbjct: 125 ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGGFTANL 184

Query: 188 FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247  
 + T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG  
 Sbjct: 185 LIVTTDVLLSGISTEAAAAFNPQMHVSVIDNWYFMASVVVLTIVGGLITDKIIEPRLGQ 244

15 Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRHPETGLVA 307  
 +Q + ++ + + S GL AGVV + A +A ++P +GILR P V  
 Sbjct: 245 WQGNSEKLTQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298

Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVNVNAMESMSTLGLYLXXXXXXXXXX 367  
 SPF+K IV I L F + + YG TR++R + ++ + M E M + ++  
 Sbjct: 299 PSPFIKGIVPLIILFFVVSILAYGIATRTIRRRQADLPHLMIEPMKEMAGFIVMVFPPLAQF 358

20 Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQWAVTAPIF 427  
 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF  
 Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418

25 Query: 428 VPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTILISMMLP 487  
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP  
 Sbjct: 419 VPMFLLGFHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPDALGTYYSLVLP 478

Query: 488 YSAFFLIAWIALFCIWVFLGLPVGPG 514  
 Y FL+ W+ + W +++GLP+GPG  
 Sbjct: 479 YPLIFLVVLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 17

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 141):

35 1 ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA  
 51 GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTTC  
 101 ATGGCATTTC GGTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA  
 151 ATgGCTTCGC GCAGTGCCTC TATACCGTA TTTTCAGCAA CGGAAATGCG  
 201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTTgTT

5  
251 CTTCAGACGG CAGCAGGTCG GTTTTGTGTG ACACCTTgAT GCACGGAaTA  
301 TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG  
351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT  
401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCgGC  
451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCCGG  
501 ACT..

This corresponds to the amino acid sequence (SEQ ID NO: 142; ORF14):

10  
1 ..TAGAAGXXVF VFVTD SQVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA  
51 MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI  
101 SPAWISCSTF STSSICPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG  
151 RXLTNPTVSV RIMLHSG..

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 (SEQ ID NO: 142) shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) (SEQ ID NO: 144) from strain A of *N. meningitidis*:

20  
orf14.pep  
orf14a  
150 160 170 180 190 200  
25  
orf14.pep  
orf14a  
210 220 230 240 250 260  
30  
orf14.pep  
orf14a  
270 280 290 300 310 320  
35  
orf14.pep  
orf14a  
330 340 350 360 370 380

The complete length ORF14a nucleotide sequence (SEQ ID NO: 143) is:

40  
1 ATGGAGGATT TGCAGGAAAT CGGGTTCGAT GTCGCCGCCG TAAAGGTAGG  
51 TCGGCAGCGC GAACATCATC GTCTGCATCA TCCCCAGCCC GGCAACGGCG  
101 AGGCGGACGA TGTATTGTTT GCGTTCCTTT TGGTTGGCGG CTTCGATTTT  
151 TTGCGCGTCA TAGGGTGCGG CCGTGTAGCC TATCTGCCTG ATTTTCAACA  
201 GAATGTCGGA AAGGCGGATT TTGCCGTCGT CCCAGACGAC GCGGCAGCGG  
45 251 TCGTGCTGT AATTGAGGTC GATGCGGACG ATGCCGTCTG TACGCAAAAG  
301 CTGCTGTTCG ATCAGCCAGA CGCAGGCGGC GCAGGTGATG CCGCCGAGCA



5  
10  
15

```

351 TTAAACCGC CTCGCGCGTG CCGCCGTGGG TTTCCACAAA GTCGGACTGG
401 ACTTCGGGCA GGTCGTACAG GCGGATTGG TCGAGGATTT CTTGGGGCGG
451 CAGCTCGGTT TTTTGC GCGT CCGCGGTGCG TTGTTTGTA TAACTGCCCA
501 AGCCCGCGTC AATAATGCTT TGTGCGACTG CCTGACAACC GGCGCAGCAG
551 GTTTCGCGGT CTTCGTTTTT GTAACGGACG GTCAGATGCA GGTTTTCGGG
601 AACGTCCAGC CCGCAGTGGA AACAGGTTTT TTTTCATGGCA TTTTCGTTTTT
651 GTCTGTGTTT GGTGCGGCGG CACAATACTC GGCAATGGCT TCGCGCAGTG
701 CGTCTATACC GGTATTTTCA GCAACGGAAA TGCGGACGGC GGCAATTTTT
751 CCCGCAGCGT CGCGCCATAT GCCCGTGT TTGTTCTCAG ACGGCAGCAG
801 GTCGGTTTTG TTGTACACCT TGATGCACGG AATATCGCCG GCATGGATTT
851 CTTGCAGTAC GTTTTCCACG TCTTCAATCT GCTGTCCGCT GTTCGGAGCG
901 GCGGCATCGA CGACGTGCAG CAGCACATCG GCTTGCGCGG TTTCTTCCAG
951 CGTGGCGGAA AAGGCGGAAA TCAGTTTGTG CGGCAGATCG CTGACGAATC
1001 CGACGGTATC GGTCAAGATA ATGCTGCATT CGGGACTGAT GTACAGCCGC
1051 CGCGCCGTCG TGTCGAGTGT GCGGAAAAGC TGGTCTTTCG CATATATGCC
1101 CGACTTGGTC AGCCGGTTGA ACAGACTGGA TTTGCCGACA TTGGTATAG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 144):

20  
25

```

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHPQP GNGEADDVLF AFFLVGGFDF
51 LRVIGCGGVA YLPDFQQNVG KADFAVVPDD AAARAVIEV DADDAVCTQK
101 LLFDQPDAGG AGDAAEH*NR LARAAGVGHK VGLDFGQVVQ ADLVEDFLGR
151 QLGFRLRVGA LFVITAQARV NNALCDCLTT GAAGFAVFVF VTDGQMVFVG
201 NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
251 PAASRHMPVF CSSDGSRSVL LYTLMHGISP AWISCSTFST SSICCPLFGA
301 AASTTCSSTS ACAVSSSVAE KAEISLCGRS LTNPTVSVRI MLHSGLMYSR
351 RAVVSSVAKS WSFAYMPDLV SRLNRLDLPT LV*

```

It should be noted that this sequence includes a stop codon at position 118.

#### Homology with a predicted ORF from *N.gonorrhoeae*

30 ORF14 (SEQ ID NO: 142) shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) (SEQ ID NO: 146) from *N. gonorrhoeae*:

35  
40

```

orf14.pep                                TAGAAGXXVFVFTDSQVEVFGNIQTAVET    30
orf14ng      GRQFGFFRVGGASFVITAQAGIDDALCDCLTADAAGFAVFVADGQMVFVGNVQPAVET    208

orf14.pep      GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS    90
orf14ng      GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS    268

orf14.pep      VLLYTLMHGISP AWISCSTFSTSSICCPLFGAAASTTCSSTSACAVSSSVAE KAEISLCG    150
orf14ng      VLLYTLMHGISWAWISCSTFSTSSICCPLFRAAASTTCSSTSACTVSSKVAE KAEISLCG    328

orf14.pep      RXLTNPTVSVRIMLHSG                                167
orf14ng      RSLTNPTVSVRIMLHAGLMYSRRVVSRAKSWSFAYMPDLVSRLNRLDLPTLV    382

```

45 The complete length ORF14ng nucleotide sequence (SEQ ID NO: 145) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 146):

```

1  MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
51  LRVIGCGGVA CLPDFQQNVG EADFAVVPDD AA AVR AVIEV DADDAVCAQK
101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMVFVG
5  201 NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
251 PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPFLRA
301 AASTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
351 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*

```

10 Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 18

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 147):

```

15 1  ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
51  GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTTGA
101 TGCCGAAGCTC GGGCAGCTTC GGTTCGGCT ATGCGTCGCT GGCGGCTTTC
151 TCGTTCGGCG CGCTGATGAT TGCCTGTGTA GACGTGTCGT CAAATATGGC
20 201 GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
301 GTGGCGGCGA TTCTGCCGTT TGTGTTTTCG TATATCGGTT TGGCGAACAC
351 CGCCGANAAA GCGTGTGTC CGCAGACCGT GGTGTCGGCG TTTTATGTGG
401 GTGCGCGGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCCGC
25 501 GAATCAGGAA AAAGCCAAC TGGATCGCACT CTTAAAAA.CC GCGC..

```

This corresponds to the amino acid sequence (SEQ ID NO: 148; ORF16):

```

30 1  ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
51  SFGALMIALL DVSSNMAMQP FKMVGDGMVN EEQKXYAYGI QSFLANTGAV
101 VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
151 EYXPETYARY HGIDVAANQE KANWIALLKX A..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 149):

```

35 1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
101 CCTTTACCCT GCAAAGCTCG CAAATGAGCC GCATTTTTC AACGCTAGGC
151 GCAGACCCGC ACAATTTGGG CTGGTTTTC ATCCTGCCGC CGCTGCGGGG
201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
40 251 CGCGTTTGGG CGGCCCGCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
351 CTATGCGTCG CTGGCGGCTT TGTCGTTTCG CGCGCTGATG ATTGCGCTGT
401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
45 501 CTTAGCAAAT ACGGGCGCGG TCGTGCGGCG GATTCTGCCG TTTGTGTTTG
551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
601 GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGTTGA TTACCAGCGC
651 GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAACCC TACGCCCGTT
701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA

```

-173-

5  
10  
15  
20  
25

```

751 CTCTTGAAAA CCGCGCCTAA GCGGTTTTGG ACGGTTACTT TGGTGCAATT
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
1001 CGGGTATTAT CCGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
1101 CATCGCTTGG GCGGGCATT TCACTTATCC GCTGACGATT GTGACCAACG
1151 CCTTGTCTGG CAAGCATATG GGCACCTACT TGGGCTTGTT TAACGGCTCT
1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGTCGCTCC
1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTCTGA TTAAAGAAAC ACACGCGGGG
1351 GTTTGA

```

15 This corresponds to the amino acid sequence (SEQ ID NO: 150; ORF16-1):

20  
25

```

1 MSEYTPQTAQ QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
51 ADPHNLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGGRR LPYLLYGTLI
101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
151 DMVNEEQKGY AYGISQFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201 VVVAFYVGAA LLVITSAFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYY
301 EAGNWYGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNCS
401 ICMPQIVASL LSFVLPFMLG GLQATMPLVG GVVLLLGAFS VFLIKETHGG
451 V*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF16 (SEQ ID NO: 148) shows 96.7% identity over a 181aa overlap with an ORF (ORF16a)  
(SEQ ID NO: 152) from strain A of *N. meningitidis*:

35  
40  
45

```

                                10      20      30
orf16.pep                                GHYSDRTWKPRLXGRRLPYLLYGTLIAVIV
                                |||||
orf16a      IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGGRRLPYLLYGTLIAVIV
              50      60      70      80      90      100

              40      50      60      70      80      90
orf16.pep      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI
              |||||
orf16a      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGI
              110     120     130     140     150     160

              100     110     120     130     140     150
orf16.pep      QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK
              |||||
orf16a      QSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLVITSAFTIFKVK
              170     180     190     200     210     220

              160     170     180
orf16.pep      EYXPETYARYHGIDVAANQEKANWIALLKXA
              |||||

```

orf16a EYNPETYARYHGIDVAANQEKANWIELLKTA PKAFWTVTLVQFFCWFAPQYMWYTYSAGAI  
 230 240 250 260 270 280

orf16a AENVHHTDASSVG YQEAGN WYGVLAAVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGA  
 290 300 310 320 330 340

The complete length ORF16a nucleotide sequence (SEQ ID NO: 151) is:

1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC  
 51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG  
 101 CCTTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT  
 151 GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG  
 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC  
 251 CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT  
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG  
 351 CTATGCGTCG CTGGCGGCTT TGTCGTTCCG CGCGCTGATG ATTGCGCTGT  
 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC  
 451 GACATGGTCA ACAGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT  
 501 CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG  
 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGACAGACC  
 601 GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC  
 651 GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGGAAACC TACGCCCGTT  
 701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA  
 751 CTCTTGAAAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT  
 801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA  
 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTTATCAG  
 901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC  
 951 GGTGATTTGT TCGTTTGAT TGGCGAAAGT GCCGAATAAA TACCATAAGG  
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT  
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG  
 1101 CATCGCTTGG GCGGGCATT TCACTTATCC GCTGACGATT GTGACCAACG  
 1151 CCTTGTCGGG CAAGCATATG GGCATTACT TGGGCCTGTT TAACGGCTCT  
 1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC  
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCCTCGTCC  
 1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTGA TTAAGAAAC ACACGGCGGG  
 1351 GTTTGA

This encodes a protein having amino acid sequence (SEQ ID NO: 152):

1 MSEYTPQTAK QGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG  
 51 ADPHSLGWFF ILPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI  
 101 AVIVMILMPN SGSFGFGYAS LAALSFGLM IALLDVSSNM AMQPFKMMVG  
 151 DMVNEEQKY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT  
 201 VVFAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE  
 251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTYS AGAIAENVWH TTDASSVG YQ  
 301 EAGN WYGVL AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV  
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLNGS  
 401 ICMPQIVASL LSFVLFPM LG LQATMFLVG GVVLLLGA FS VFLIKETHGG  
 451 V\*

ORF16a (SEQ ID NO: 152) and ORF16-1 (SEQ ID NO: 150) show 99.6% identity in 451 aa overlap:

50 orf16a.pep 10 20 30 40 50 60  
 MSEYTPQTAKQGLPALAKSTIWMLSFGLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF  
 |||||

5	orf16-1	MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF	10	20	30	40	50	60
	orf16a.pep	ILPPLAGMLVQPIVGHYSDRTWKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS	70	80	90	100	110	120
10	orf16-1	ILPPLAGMLVQPIVGHYSDRTWKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS	70	80	90	100	110	120
	orf16a.pep	LAALSFGALMIALLDVSSNMAMQPFKMMVGDMDVNEEQGYAYGIQSFLANTGAVVAAILP	130	140	150	160	170	180
15	orf16-1	LAALSFGALMIALLDVSSNMAMQPFKMMVGDMDVNEEQGYAYGIQSFLANTGAVVAAILP	130	140	150	160	170	180
	orf16a.pep	FVFAYIGLANTAEGVVPQTVVFAFYVGAALLVITSAFTIFKVKEYNPETYARYHGIDVA	190	200	210	220	230	240
20	orf16-1	FVFAYIGLANTAEGVVPQTVVFAFYVGAALLVITSAFTIFKVKEYNPETYARYHGIDVA	190	200	210	220	230	240
	orf16a.pep	ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI AENVWHTTDASSVGYQ	250	260	270	280	290	300
25	orf16-1	ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI AENVWHTTDASSVGYQ	250	260	270	280	290	300
	orf16a.pep	EAGNWyGVLA AVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV	310	320	330	340	350	360
30	orf16-1	EAGNWyGVLA AVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV	310	320	330	340	350	360
	orf16a.pep	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNCSICMPQIVASLLSFVLPMLG	370	380	390	400	410	420
35	orf16-1	LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNCSICMPQIVASLLSFVLPMLG	370	380	390	400	410	420
	orf16a.pep	GLQATMFLVGGVVLLLGAFSVFLIKETHGGVX	430	440	450			
40	orf16-1	GLQATMFLVGGVVLLLGAFSVFLIKETHGGVX	430	440	450			

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 (SEQ ID NO: 148) shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) (SEQ ID NO: 154) from *N. gonorrhoeae*:

45	orf16.pep	GHYSDRTWKPRXLGRRRLPYLLYGTLIAVIV	30
	orf16ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYYSRTWKPRLGGRRLPYLLYGTLIAVIV	131
45	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMDVNEEQXYAYGI	90
	orf16ng		

orf16ng	MILMPNSGSFSGFYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNVNEEQKSYAYGI	191
orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK	150
orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEKGVVPQTVVVAFYVGAALLIITSAFTISKVK	251
5 orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVPVQFFCWFAFRYMWTYSAGAI	311

The complete length ORF16ng nucleotide sequence (SEQ ID NO: 153) is:

10	1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
	51	TACTTTTCAA	ATCAAAAAAA	AGGATTACT	TTATGTCGGA	ATATACGCCT
	101	CAAACAGCAA	AACAAGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTGGAT
	151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
	201	CGCAGATGAG	CCGCATTTT	CAAACGCTAG	GCGCAGACCC	GCACAATTTG
15	251	GGCTGGTTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG	TTCAGCCGAT
	301	AGTGGCTACT	ACTCAGACCG	CACTTGGAAG	CCGCGCTTGG	GCGGCCGCCG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC	GTGATGATTT
	401	TGATGCCGAA	CTCGGGCAGC	TTCGGTTTCG	GCTATGCGTC	GCTGGCGGCC
	451	TTGTCGTTTCG	GCGCGCTGAT	GATTGCGCTG	TTGGACGTGT	CGTCGAATAT
20	501	GGCGATGCAG	CCGTTTAAGA	TGATGGTCGG	CGATATGGTC	AACGAGGAGC
	551	AGAAAAGCTA	CGCCTACGGG	ATTCAAAGTT	TCTTAGCGAA	TACGGACGCG
	601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTC	GCGTATATCG	GTTTGGCGAA
	651	CTAGTCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA	GCATTCTATG
	701	TGGGTGCGGC	GTTACTGATT	ATTACCAGTG	CGTTCACAAT	CTCCAAAGTC
25	751	AAAGAATACG	ACCCGGAAAC	CTACGCCCGT	TACCACGCA	TCGATGTCGC
	801	CGCGAATCAG	GAAAAAGCCA	ACTGGTTCGA	ACTCTTAAAA	ACCGCGCCTA
	851	AAGTGTTTTG	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG	GTTTCGCCTTC
	901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA	ACGTCTGGCA
	951	CACTACCGAT	CGCTCTCCG	TAGGCCATCA	GGAGGCGGGC	AACCGGTACG
30	1001	GCGTTTTTGGC	GGCGGTGTAG			

This encodes a protein having amino acid sequence (SEQ ID NO: 154):

	1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLYVG	IYASNSKTRF	ARAGKKHDL
	51	VELRLSRRSD	GLYPAKLADE	PHFSNARRRP	AQFGLVFHPA	AAGGDAGSAD
35	101	SGYYSDRTWK	PRLGGRRLPY	LLYGTLIAVI	VMILMPNSGS	FGFGYASLAA
	151	LSFGALMIAL	LDVSSNMAMQ	PFKMMVGDV	NEEQKSYAYG	IQSFLANTDA
	201	VVAAILPFVF	AYIGLANTAE	KGVVPQTVVV	AFYVGAALLI	ITSAFTISKV
	251	KEYDPETYAR	YHGIDVAANQ	EKANWFELLK	TAPKVFWTVT	PVQFFCWFAF
40	301	RYMWTYSAGA	IAENVWHTTD	ASSVGHQEAG	NRYGVLAHV*	

ORF16ng (SEQ ID NO: 154) and ORF16-1 (SEQ ID NO: 150) show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
45	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYS					
	orf16ng	DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGDAGSADSGYYS					
		50	60	70	80	90	100
		90	100	110	120	130	140
50	orf16-1.pep	WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFSGFYASLAALSFGALMIALLDVSSNMA					

	orf16ng	WKPRLGGRRLPYLLYGT	LI	AV	IV	MIL	MP	NS	GS	FG	FG	YAS	LA	ALS	SF	GAL	MIAL	LD	VSS	NMA
		110	120	130	140	150	160													
5	orf16-1.pep	150	160	170	180	190	200													
	orf16ng	170	180	190	200	210	220													
10	orf16-1.pep	210	220	230	240	250	260													
	orf16ng	230	240	250	260	270	280													
15	orf16-1.pep	270	280	290	300	310	320													
	orf16ng	290	300	310	320	330	340													

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 19

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 155):

25	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
30	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCCGCG	AA.NTGACGG
	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACCTG	GATCGNCTGG
	351	CAGCCAGAAT...				

This corresponds to the amino acid sequence (SEQ ID NO: 156; ORF28):

35	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNPN	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFFVNPEDSA	XXTGILXAGL	DKPFIQVXDT
	101	PSYXCHQALP	VKLGSXGSQN...			

Further work revealed the complete nucleotide sequence (SEQ ID NO: 157):

40	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCCGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC

301 CCGAGCTATG CTCGCCACCA AGCCCTGCCG GTCAAACCTCG AATCGCCTGG  
 351 CAGCCAGAAT TTCAGTACCG AAGGCCTTTG CCTGCGCTAC GATACCGACA  
 401 AGCCTGCCGA CATCGCCAAG CTGAAACAGC TCGGGTTTGA AGCGGTCAAA  
 451 CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA  
 501 CTACGCCACA CCGCAAAAAC TGAACGCCGA TTACCATTTT GAGCAAAGTG  
 551 TGCCTGCCGA TATTTATTAC ACGGTTACTG AAGAACATAC CGACAAATCC  
 601 AAGCTGTTTG CAAATATCTT ATATACGCCC CCCTTTTGA TACTGGATGC  
 651 GCGGGCGCG GTACTGGCCT TGCCTGCGGC GGCTCTGGGT GCGGTCGTGG  
 701 ATGCCGCCCG CAAATGA

This corresponds to the amino acid sequence (SEQ ID NO: 158; ORF28-1):

1 MLFRKTTAAV LAATLMLNGC TLMLWGMNPN VSETITRKHV DKDQIRAFGV  
 51 VAEDNAQLEK GSLVMMGGKY WFFVNPEDSA KLTGILKAGL DKPFQIVEDT  
 101 PSYARHQALP VKLESPGSQN FSTEGCLRLY DTDKPADIAK LKQLGFPAVK  
 151 LDNRTIYTRC VSAKGKYYAT PQKLNADYHF EQSVPADIYY TVTEEHTDKS  
 201 KLFANILYTP PFLILDAAGA VLALPAAALG AVVDAARK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 (SEQ ID NO: 156) shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) (SEQ ID NO: 160) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf28.pep		MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK					
		:     :     :     :     :     :     :					
orf28a		MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf28.pep		GSLVMMGGKYWFFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSON					
		:     :     :     :     :     :     :					
orf28a		GSLVMMGGKYWFFVNPEDSAKLTGILKAGLDKQFQMVPEPNRFA-YQALPVKLESPASQN					
		70	80	90	100	110	
orf28a		FSTEGCLRLYDTRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		120	130	140	150	160	170

The complete length ORF28a nucleotide sequence (SEQ ID NO: 159) is:

1 ATGTTGTTCC GTAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATGTT  
 51 GAACGGCTGT ACGGTAATGA TGTGGGGTAT GAACAGCCCG TTCAGCGAAA  
 101 CGACCGCCCG CAAACACGTT GACAAGGACC AAATCCGCGC CTTCCGTGTG  
 151 GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG  
 201 CGGGAAATAC TGGTTCGTCG TCAATCCTGA AGATTCGGCG AAGCTGACGG  
 251 GCATTTTGAA GGCCGGGTTG GACAAGCAGT TTCAAATGGT TGAGCCCAAC  
 301 CCGCGCTTTG CCTACCAAGC CCTGCCGGTC AAATCGAAT CGCCCGCCAG  
 351 CCAGAATTC AGTACCGAAG GCCTTTGCCT GCGCTACGAT ACCGACAGAC  
 401 CTGCCGACAT CGCCAAGCTG AAACAGCTTG AGTTTGAAGC GGTGGAATC  
 451 GACAATCGGA CCATTTACAC GCGCTGCGTC TCCGCCAAAG GCAAACTACTA  
 501 CGCCACACCG CAAAACCTGA ACGCCGATTA TCATTTTGAG CAAAGTGTGC  
 551 CTGCCGATAT TTATTACACG GTTACGAAAA AACATACCGA CAAATCCAAG  
 601 TTGTTTGAAA ATATTGCATA TACGCCACC ACGTTGATAC TGGATGCGGT



651 GGGCGCGGTG CTGGCCTTGC CTGTCGCGGC GTTGATTGCA GCCACGAATT  
701 CCTCAGACAA ATGA

This encodes a protein having amino acid sequence (SEQ ID NO: 160):

5           1   MLFRKTTAAV LAATLMLNGC TVMMWGMNSP FSETTARKHV DKDQIRAFGV  
          51   VAEDNAQLEK GSLVMMGGKY WFVVPEDSA KLTGILKAGL DKQFQMVEPN  
         101   PRFAYQALPV KLESPASQNF STEGLCLRYD TDRPADIACL KQLEFEAVEL  
         151   DNRTIYTRCV SAKGKYATP QKLNADYHFE QSVPADIIYT VTKKHTDKSK  
         201   LFENIAYTPT TLILDAVGAV LALPVAALIA ATNSSDK\*

ORF28a (SEQ ID NO: 160) and ORF28-1 (SEQ ID NO: 158) show 86.1% identity in 238 aa overlap:

		10	20	30	40	50	60
15	orf28a.pep	MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
	orf28-1	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
20	orf28a.pep	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKQFQMVEPNPRFA-YQALPVKLESPASQN					
	orf28-1	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN					
		70	80	90	100	110	119
25	orf28a.pep	FSTEGLCLRYDTDRPADIACLKQLEFEAVELDNRTIYTRCVSAKGKYATPQKLNADYHF					
	orf28-1	FSTEGLCLRYDTDKPADIACLKQLGFEAVKLDNRTIYTRCVSAKGKYATPQKLNADYHF					
		120	130	140	150	160	170
30	orf28a.pep	EQSVPADIIYTVTKKHTDKSKLFENIAYTPTTLILDAVGAVLALPVAALIAATNSSDKX					
	orf28-1	EQSVPADIIYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
		180	190	200	210	220	230

Homology with a predicted ORF from *N.gonorrhoeae*

ORF28 (SEQ ID NO: 156) shows 84.2% identity over a 120aa overlap with a predicted ORF  
(ORF28.ng) (SEQ ID NO: 162) from *N. gonorrhoeae*:

	orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKDQIRAFGVVAEDNAQLEK	60
40	orf28.pep	GSLVMMGGKYWFVVPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGXGSGQN	120
	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN	120

The complete length ORF28ng nucleotide sequence (SEQ ID NO: 161) is

-180-

5  
10  
15

```

1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATACT
51  GAACGGCTGT ACGATGATGT TGCGGGGGAT GAACAACCCG GTCAGCCAAA
101 CAATCACCCG CAAACACGTT GACAAAGACC AAATCCGCGC CTTCGGTGTG
151 GTTGCCGAAG ACAATGCCCA ATTGGAAGAG GGCAGCCTGG TGATGATGGG
201 CGGGAAATAC TGGTTCGCCG TCAATCCCGA AGATTCGGCG AAGCTGACGG
251 GCCTTTTGAA GGCCGGGTTG GACAAGCCCT TCCAAATAGT TGAGGATACC
301 CCGAGCTATG CCCGCCACCA AGCCCTGCCG GTCAAATTCG AAGCGCCCGG
351 CAGCCAGAAT TTCAGTACCG GAGGTCTTTG CCTGCGCTAT GATACCGGCA
401 GACCTGACGA CATCGCCAAG CTGAAACAGC TTGAGTTTAA AGCGGTCAAA
451 CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA
501 CTACGCCACG CCGCAAAAAC TGAACGCCGA TTATCATTTT GAGCAAAGTG
551 TGCCCGCCGA TATTTATTAT ACGGTTACTG AAAAACATAC CGACAAATCC
601 AAGCTGTTTG GAAATATCTT ATATACGCCC CCCTTGTTGA TATTGGATGC
651 GCGCGCCGCG GTGCTGGTCT TGCCTATGGC TCTGATTGCA GCCGCGAATT
701 CCTCAGACAA ATGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 162):

20

```

1  MLFRKTTAAV LAATLILNGC TMMLRGMNPN VSQTITRKHV DKDQIRAFGV
51  VAEDNAQLEK GSLVMMGGKY WFAVNPEDSA KLTGLLKAGL DKPFQIVEDT
101 PSYARHQALP VKFEAPGSQN FSTGGLCLRY DTGRPDIAK LKQLEFKAVK
151 LDNRTIYTRC VSAKGKYYAT PQKLNADYHF EQSVPADIYY TVTEKHTDKS
201 KLFGNILYTP PLLILDAAAA VLVLPMALIA AANSSDK*

```

25 overlap:

30

```

      10      20      30      40      50      60
orf28-1.pep MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKQIRAFGVVAEDNAQLEK
orf28ng      MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKQIRAFGVVAEDNAQLEK
      10      20      30      40      50      60

      70      80      90      100     110     120
orf28-1.pep GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN
orf28ng      GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN
      70      80      90      100     110     120

      130     140     150     160     170     180
orf28-1.pep FSTGGLCLRYDTDKPADIAPKQKLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
orf28ng      FSTGGLCLRYDTGRPDIAKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
      130     140     150     160     170     180

      190     200     210     220     230     239
orf28-1.pep EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX
orf28ng      EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAALVLVLPMALIAAANSSDKX
      190     200     210     220     230

```

45

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (SEQ ID NO: 158) (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 (SEQ ID NO: 158) is a surface-exposed protein, and that it may be a useful immunogen.

### Example 20

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 163):

```

10      1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
      51  TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
     101  CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTTCA CGGCGGTGTA
     151  GACGGCGGTT TTAATGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
     201  TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCGG
     15  251  GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
     301  ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
     351  AGAAATGCC GGTGCCGCCT CTGGT..

```

This corresponds to the amino acid sequence (SEQ ID NO: 164; ORF29):

```

20      1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHHDS KSTSDFSGGV
     51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
     101  TKTSIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 165):

```

25      1  ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
     51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
     101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
     151  TTTGGTAATG CTCGCGGCAG TGTAAAAAG CGGGTTTACG CCGTCCAGAC
     201  ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
     30  251  CAGGGTTTGA AGGTGTTATC GGTATGAAA CCCATTTTTC AGGGCACGGA
     301  CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
     351  TTTCAGCGGC GGTGTAGACG GCGGTTTTAC TGTTTACCAA CTTTCATCGAA
     401  CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
     451  GATTATCCGC CCCCCGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
     35  501  AGGAACTTCA ACAAAAACAA AGACTAATAT TGTCCCTCAA GCCCATTTT
     551  CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
     601  CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
     651  TTGGTGGGCT AACCGTATGG ATGATGTTTC CGGCATCGTC CAAGGTGCGG
     701  TTAATCCTTT TTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
     40  751  GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
     801  AGGTATTAAT GATTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
     851  CGAGCCTATT ACAGGACAGT GCTTTTTCGG TAAAAGACGG TATCAACTCT
     901  GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAACAGCTA CTGCCCAAAC
     951  TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTCGAGA GGTAAAAAAG

```



1 ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC  
 51 GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC  
 101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG  
 151 TTTGGTAATG CTCGCGGCAG TGTAAAAAAT CGGGTTTACG CCGTCCAAAC  
 5 201 ATTTGATGCA ACTGCGGTCTG GCCCCATACT GCCTATTACA CACGAACGGA  
 251 CAGGATTTGA AGGCATTATC GGTATGAAA CCCATTTTTC AGGACATGGA  
 301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACCTTCTGA  
 351 TTTCAGCGGC GGCCTAGACG GTGGTTTAC CGTTTACCAA CTTTCATCGGA  
 401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC  
 10 451 GATTATCCGC CCCCCGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA  
 501 AGGAACCTCA ACAAAAAACA AGAGTAATAT TGTTCCTCGA GCCCCATTTT  
 551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTCTGG TTTTTCAGC  
 601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA  
 651 TTGGTGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG  
 15 701 TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA  
 751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA  
 801 AGGTATNAAT CATTTAGGAA ANTTAAGTCC CGAAGCACAA CTTGCGGCTG  
 851 CAACCGCATT ACAAGACAGT GCTTTTGCAG TAAAAGACGG TATCAATTCC  
 901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACGCAA CAGCCCAAAC  
 20 951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAAAAAAG  
 1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TAAAAAATAC NGGCTATAAN  
 1051 ACACCTGCTG TTCGCACCAT GCATACTTTG GATGGGGAAA TGGCCGGTGG  
 1101 GAATAGACCG CTTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA  
 1151 CACAACCGTC TTTACAAGCG CAACTAATTG GAGAACAAAT TANNNNNGGG  
 25 1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTTAAA  
 1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAT ATTGTAGCC  
 1301 ATCCANCAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT  
 1351 NATAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG  
 30 1401 TACAGCATTT AGACCAACAT CAGGTAAAA ATATTATGAT GATTATAG

This encodes a protein having amino acid sequence (SEQ ID NO: 168):

1 MNXPIQKFMM LFAAAISXLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL  
 51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG  
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS  
 35 151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS  
 201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT  
 251 DSAVSPVTD TAAQTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGIN  
 301 ARQWADAHPN ITATAQTALA VAXAATTVWG GKKVELNPTK WDWVKNTGYX  
 351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEQIXXG  
 40 401 HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPXNMKE LPRGRTAYWD  
 451 KKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL\*

ORF29a (SEQ ID NO: 168) and ORF29-1 (SEQ ID NO: 166) show 90.1% identity in 385 aa overlap:

45  
 orf29a.pep      10      20      30      40      50      60  
 MNXPIQKFMM LFAAAISXLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN  
 orf29-1      MNLP IQKFMM LFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK  
 10      20      30      40      50      60  
 50  
 orf29a.pep      70      80      90      100      110      120  
 RVYAVQTFDATAVGPILPITHERTGFEGIIIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG  
 orf29-1      RVYAVQTFDATAVSPVLPITHERTGFEGVIGYETHFSGHGHEVHSPFDHDSKSTSDFSG  
 70      80      90      100      110      120

5	orf29a.pep	130	140	150	160	170	180
	orf29-1	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS	DYPPPGGARDIYXXYVKG	TSTKTKSNIVPR			
10	orf29a.pep	190	200	210	220	230	240
	orf29-1	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMD	DIRGIVQGAVNPFLMG				
15	orf29a.pep	250	260	270	280	290	300
	orf29-1	FQGVGIGAITDSAVSPVTD	TAAQQT	LQGXNHLGXL	SPEAQLAAATALQDS	FAVKD	GINS
20	orf29a.pep	310	320	330	340	350	360
	orf29-1	ARQWADAHPNITATAQTALAVAXAATT	VWGGKKVELNPTKWDVVKNTGYXTPAVRTMHTL				
25	orf29a.pep	370	380	390	400	410	420
	orf29-1	DGEMAGGNRPPKSITSNSKADASTQPSLQAQLIGE	QIXXGHAYNKHVIRQQEFTDLNINS				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 (SEQ ID NO: 164) shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) (SEQ ID NO: 170) from *N. gonorrhoeae*:

30	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE	102
35	orf29.pep	VHSPFDHHD	90
	orf29ng	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGYPPPGGARDIY	162
35	orf29.pep	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANR	222

The complete length ORF29ng nucleotide sequence (SEQ ID NO: 169) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 170):

```

1  MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPPGGKYHL
51  FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGG
151 GYPPPGGARD IYSYHIKGT TKTINTVPQ APFSDRWLKE NAGAASGFLS

```

5  
 201 RADEAGKLIW ENDPDKNWRA NRMDDIRGIV QGAVNPFLTG FQGLGVGAIT  
 251 DSAVSPVITYA AARKTLQGIH NLGNLSPEAQ LAAATALQDS AFAVKDSINS  
 301 ARQWADAHPN ITATAQTALA VTEAATTVWG GKKVELNPAK WDWVKNTGYK  
 351 KPAARHMQTV DGEMAGGNKP LESKNTVTTN NFFENTGYTE KVLQASNGD  
 401 YHGFPQSVDA FSENGTVIQI VGGDNIVRHK LYIPGSYK GK DGNFEYIREA  
 451 DGKINHRLFV PNQQLPEK\*

In a second experiment, the following DNA sequence (SEQ ID NO: 171) was identified:

10  
 1 atgAATTTGC CTATTCAAAA ATTCATGATG ctgttggcAg cggcaatatac  
 51 gatgctGCat ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC  
 101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGCAA ATACCATCTG  
 151 TTTGGTAATG CTCGCGGCAG TGTAAAAAAT CGGGTTTGC GCGTCCAAAC  
 201 ATTTGATGCA ACTGCGGTCG GCCCATACT GCCTATTACA CACGAACGGA  
 251 CAGGATTTGA AGGTGTTATC GGCATGAAA CCCATTTTTC AGGACACGGA  
 15  
 301 CACGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA  
 351 TTTACGCGGC GCGGTAGACG GCGGTTTTAC CGTTTACCAA CTTCATCGGA  
 401 CAGGGTCGGA AATACATCCC GCAGACGGAT ATGACGGGCC TCAAGGCGGC  
 451 GGTATATCCG AACCACAAGG GGCAAGGGAT ATATACAGCT ACCATATCAA  
 501 AGGAACTTCA ACCAAAAACA AGATAAACAC TGTCCGCAA GCCCCTTTTT  
 20  
 551 CAGACCGCTG GCTAAAGAA AATGCCGGTG CCGCTTCCG TTTTCTCAGC  
 601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAACGACC CCGATAAAAA  
 651 TTGGCGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG  
 701 TTAATCCTTT TTTAACGGGT TTTCAAGGGG TAGGGATTGG GGCAATTACA  
 751 GACAGTGCGG TAAGCCCGGT CACAGATACA GCCGCTCAGC AGACTCTACA  
 25  
 801 AGGTATTAAT GATTTAGGAA ATTTAAGTCC GGAAGCACAA CTGCCGCCG  
 851 CGAGCTATT ACAGGACAGT GCCTTTGCGG TAAAAGACGG CATCAATTCC  
 901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACAGCAA CAGCCCAAAC  
 951 TGCCCTTGCC GTAGCAGAG CCGCAGGTAC GGTTTGGCGC GGTAATAAAG  
 1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TTAATAATAC CGGCTATAAA  
 30  
 1051 AAACCTGCTG CCCGCCATAT GCAGACTGTA GATGGGGAGA TGGCAGGGGG  
 1101 GAATAGACCG CCTAAATCTA TAACGTCGGA AGGAAAAGCT AATGCTGCAA  
 1151 CCTATCCTAA GTTGGTTAAT CAGCTAAATG AGCAAAACTT AAATAACATT  
 1201 GCGGCTCAAG ATCCAAGATT GAGTCTAGCT ATTCATGAGG GTAAAAA  
 1251 TTTTCCAATA GGAAGTCAA CTTATGAAGA GGCAGATAGA CTAGGTAAAA  
 35  
 1301 TTTGGGTTGG TGAGGTGCA AGACAACTA GTGGAGGCGG ATGGTTAAGT  
 1351 AGAGTGGCA CTCGACAATA TCGGCCACCA ACAGAAAAA AATCACAATT  
 1401 TGCAACTACA GGTATTCAAG CAAATTTTGA AACTTATACT ATTGATTCAA  
 1451 ATGAAAAAAG AAATAAAATT AAAAAATGGAC ATTTAAATAT TAGGTAA

40 This encodes a protein having amino acid sequence (SEQ ID NO: 172; ORF29ng-1):

45  
 1 MNLPIQKFMM LLAAAI SMLH IPISHANGLD ARLRDMQAK HYEPPGKYHL  
 51 FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG  
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG  
 151 GYPEPQGARD IYSYHIKGT TTKKINTVPQ APFSDRWLKE NAGAASGFLS  
 201 RADEAGKLIW ENDPDKNWRA NRMDDIRGIV QGAVNPFLTG FQGVGIGAIT  
 251 DSAVSPVTD T AAQTLQGIN DLGNLSPEAQ LAAASLLQDS AFAVKDGINS  
 301 ARQWADAHPN ITATAQTALA VAEAGTVWR GKKVELNPTK WDWVKNTGYK  
 351 KPAARHMQTV DGEMAGGNRP PKSITSEGKA NAATYPKLVN QLNEQNLNNI  
 401 AAQDPRLSLA IHEGKNFPI GTATYEEADR LGKIWVGEGA RQTSGGGWLS  
 50  
 451 RDGTRQYRPP TEKKSQFATT GIQANFETYT IDSNEKRNKI KNHGLNIR\*

ORF29ng-1 (SEQ ID NO: 172) and ORF29-1 (SEQ ID NO: 166) show 86.0% identity in 401 aa overlap:

-186-

		10	20	30	40	50	60
	orf29ng-1.pep	MNLPIQKFMMLLAAAISMLHIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN					
5	orf29-1	MNLPIQKFMMLFAAAISLLQIPIISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK					
		10	20	30	40	50	60
	orf29ng-1.pep	70	80	90	100	110	120
	orf29-1	RVCAVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHEVHSPFDNHDSTSDFSG					
10		70	80	90	100	110	120
	orf29ng-1.pep	130	140	150	160	170	180
	orf29-1	GVDGGFTVYQLHRTGSEIHPADGYDGPQGGGYPEPQGARDIYSYHIKGTSTKTKINTVPQ					
15		130	140	150	160	170	180
	orf29ng-1.pep	190	200	210	220	230	240
	orf29-1	APFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANRMDDIRGIVQGAVNPFLTG					
20		190	200	210	220	230	240
	orf29ng-1.pep	250	260	270	280	290	300
	orf29-1	FQGVGIGAITDSAVSPVTDTAQAQTLQGINDLGNLSPEAQLAAASLLQDSFAVKDGIN					
25		250	260	270	280	290	300
	orf29ng-1.pep	310	320	330	340	350	360
	orf29-1	ARQWADAHPNITATAQTALAVAEAAAGTVWRGKKVELNPTKWDVVKNTGYKKPAARHMQTV					
30		310	320	330	340	350	360
	orf29ng-1.pep	370	380	390	400	410	419
	orf29-1	DGEMAGGNRPPKSI-TSEGGKANAATYPKLVNQLNEQNLNIAAQDPRLSLAIHEGKKNFP					
35		370	380	390	400	410	420
	orf29ng-1.pep	420	430	440	450	460	470
	orf29-1	IGTATYEEADRLGKIWVGEGARQTSGGGWLSRDGTRQYRPPTEKKSQFATTGIQANFETY					
40		420	430	440	450	460	470
	orf29ng-1.pep	420	430	440	450	460	470
	orf29-1	RYTSLDGKITIIKDNENNYFRIHDNSRKQYLDNSNGNAVKTGNLQGGKQAKDYLLQQQTHIRN					
		430	440	450	460	470	480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 21



The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 173):

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAG...

```

This corresponds to the amino acid sequence (SEQ ID NO: 174; ORF30):

```

1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 175):

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
151 ATGAAGGAGA CAGAGGGGGC GTTCTTTCCA TTGGCTATCT TGGGTGGTGC
201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCCT TTCGGTAATA
401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 176; ORF30-1):

```

1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
51 MKETEGAFIP LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI
101 PGGVGAAGKV VSAFYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF30 (SEQ ID NO: 174) shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) (SEQ ID NO: 178) from strain A of *N. meningitidis*:

```

              10      20      30      40
orf30.pep    MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
              |||||
orf30a       MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
              10      20      30      40      50      60
orf30a       LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI
              70      80      90      100     110     120

```

The complete length ORF30a nucleotide sequence (SEQ ID NO: 177) is:

```

1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC

```

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

This encodes a protein having amino acid sequence (SEQ ID NO: 178):

1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE  
51 MKXTXGAFLP LXILGGAAG MWTHGFSYA TTGRPASVRD VAIAGGLGAI  
101 PGXVGAAGKV VSAFYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT  
151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F\*

ORF30a (SEQ ID NO: 178) and ORF30-1 (SEQ ID NO: 176) show 97.8% identity in 181 aa overlap:

20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 (SEQ ID NO: 174) shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) (SEQ ID NO: 180) from *N. gonorrhoeae*:

35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

The complete length ORF30ng nucleotide sequence (SEQ ID NO: 179) is

40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

5  
251 GACCAGCTTC TGTTAGAGAT GTTGCTGGCG GATTAGGCGC AATTCCTGGT  
301 GATGTAGGTG CTGCAGGAAA GGTGTTTCC TTTGCTAAAT ATGGACGTGA  
351 GATTAAAATC GGCAATAATA TGCGGATAGC CCCTTTCGGT AATAGAACAG  
401 GTCATCCTAT TGGAAAATTT CCCCATTATC ATCGTCGAGT TACGGATAAT  
451 ACGGGCAAGA CTTTGCCTGG ACAGGGAATT GGTCGTCATC GCCCTTGGGA  
501 ATCAAAATCT ACGGACAGAT CATGGAAAAA CCGCTTCTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 180):

10  
1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE  
51 MKETEGAFLLP LAILGGAAIG MWTQHGFSSYA TTGRPASVRD VAGGLGAIPG  
101 DVGAAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN  
151 TGKTLPGQGI GRHRPWESKS TDRSWKNRF\*

15 ORF30ng (SEQ ID NO: 180) and ORF30-1 (SEQ ID NO: 176) show 98.3% identity in 181 aa overlap:

20  
orf30ng.pep 10 20 30 40 50 60  
MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLLP  
orf30-1 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLLP  
10 20 30 40 50 60  
25  
orf30ng.pep 70 80 90 100 110  
LAILGGAAIGMWTQHGFSSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI  
orf30-1 LAILGGAAIGMWTQHGFSSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI  
70 80 90 100 110 120  
30  
orf30ng.pep 120 130 140 150 160 170  
KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQIGRHRPWESKSTDRSWKNR  
orf30-1 KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQIGRHRPWESKSTDRSWKNR  
130 140 150 160 170 180  
35  
orf30ng.pep 180  
FX  
||  
orf30-1 FX

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 22

40 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 181):

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT  
51 GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA  
101 GTGATTTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT

```

151 GCACCTGTTT GTg.CGTTaC AAATATCTTT TCTTTTCTT TATTGGGCTT
201 TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
251 TT..

```

5 This corresponds to the amino acid sequence (SEQ ID NO: 182; ORF31):

```

1 MNKTLRYVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..

```

Further work revealed a further partial nucleotide sequence (SEQ ID NO: 183):

```

10      1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
      51 GGTAGCCGTT GCTGAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
      101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
      151 GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTTAT TGGGCTTTTC
      201 TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..
15

```

This corresponds to the amino acid sequence (SEQ ID NO: 184; ORF31-1):

```

1 MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPFGTTH
51 APVCRSNIFS FSLLGFSLCL AVGTANIAFA DGI..

```

20 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 (SEQ ID NO: 182) shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) (SEQ ID NO: 186) from *N. gonorrhoeae*:

```

25      orf31.pep      MNKTLRYVIFNRKRGAVXAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF      60
      orf31ng      MNKTLRYVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAF      54

      orf31.pep      SFSLLGFSLCLAVGTXNIAFADGI      84
      orf31ng      CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV      114
30

```

The complete length ORF31ng nucleotide sequence (SEQ ID NO: 185) is:

```

      1 ATGAACAAAA CCCTCTATCG TGTGATTTTC AACCGCAAAC GCGGTGCTGT
      51 GGTAGCTGTT GCCGAAACCA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
35      101 GTGGTTCGGG CAGCGTTTAT GTGAAATCCG TTTCTTTTAT TCCTACTCAT
      151 TCCAAAGCCT TTTGTTTTTC TGCATTAGGC TTTTCTTTAT GTTTGGCTTT
      201 GGGTACGGTC AATATTGCTT TTGCTGACGG CATTATTACT GATAAAGCTG
      251 CTCCTAAAAC CCAACAAGCC ACGATTCTGC AAACAGGTaa cGGCATAACCG
      301 CAAGTCAATA TTCAAACCCC TACTTCGGCA GGGGTTTCTG TTAATCAATA
      351 TGCCAGTTT GATGTGGGTA ATCGCGGGGC GATTTTAAAC AACAGTCGCA
40      401 GCAACACCCA AACACAGCTA GGCGGTTGGA TTCAAGGCAA TCCTTGTTG
      451 ACAAGGGGCG AAGCACGTGT GGTGTGTAAC CAAATCAACA GCAGCCATCC
      501 TTCACAAC TG AATGGCTATA TTGAAGTGGG TGGACGACGT GCAGAAGTCG
      551 TTATTGCCAA TCCGGCAGGG ATTGCAGTCA ATGGTGGTGG TTTTATCAAT

```

601 GCTTCCCGTG CCACTTTGAC GACAGGCCAA CCGCAATATC AAGCAGGAGA  
651 CTTTAGCGGC TTTAAGATAA GGCAAGGCAA TGCTGTAATC GCCGGACACG  
701 GTTTGGATGC CCGTGATACC GATTTCACAC GTATTCTTGT ATGCCAACAA  
751 AATCACCTTG ATCAGTACGG CCGAACAAAGC AGGCATTTCGT AA

This encodes a protein having amino acid sequence (SEQ ID NO: 186):

1	MNKTLYRVIF	NRKRGAVVAV	AETTKREGKS	CADSGSGSVY	VKSVSFIPTH
51	SKAFCFSALG	FSLCLALGTV	NIAFADGIIT	DKAAPKTQQA	TILQTNGNIP
101	QVNIQTPTSA	GVSVNQYAQF	DVGNGRGAILE	NSRNTQTQL	GGWIQGNPWL
151	TRGEARVVVN	QINSSHPSQL	NGYIEVGRR	AEVVIANPAG	IAVNGGGFIN
201	ASRATLTTGQ	PQYQAGDFSG	FKIRQNGAVI	AGHGLDARD	DFTRILVCQQ
251	NHLDOYGRTS	RHS*			

This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein (SEQ ID NO: 1125) from *Erwinia chrysanthemi* (accession number L39897):

Orf31ng	96	GNGIPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE	154
		GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L	
HecA	45	GNGVPVVNIATPDASGLSHNRYHDFNVNDRGLILNNGTARLTPSQLGGLIQNNPNLNGRA	104
Orf31ng	155	ARVVVNQINSSHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQ	214
		A' ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTTG PQ+	
HecA	105	AAAILNEVVSPPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPQFD	164
Orf31ng	215	-AGDFSQFKIRQGNNAVIAGHGLDARDTDF	242
		AG SG +R G+ +I G.GLDA +D+	
HecA	165	AAGGLSGLDVRGGDILIDGAGLDASRSDY	193

Furthermore, ORF31ng (SEQ ID NO: 186) and ORF31-1 (SEQ ID NO: 184) show 79.5% identity in 83 aa overlap:

```

                                10      20      30      40      50      60
orf31-1.pep  MNKTLYRVIFNRKRGAVVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS
              |||||||||||||||||||||||||||||||||:|:||||| | || | :|
orf31ng      MNKTLYRVIFNRKRGAVVAETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAFC
              10      20      30      40      50

                                70      80
orf31-1.pep  FSLLGFSCLAVGTANIAFADGI
              || |||||||||:|:|||||||
orf31ng      FSALGFSCLALGTVNIAFADGIITDKAAPKTQQATILQTGNIGIPQVNIQTPTSAGVSVN
              60      70      80      90      100     110

```

On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 23

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 187):

```

1  ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
51  TTTCGCGGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
5  151 GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGC GTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..

```

This corresponds to the amino acid sequence (SEQ ID NO: 188; ORF32):

```

1  MNTPPFVCWI FCKVIDNFGD IGVSRLARV LHRELGWQVH LWTDDVSALR
10 51  ALCPDLPDVP CVHQDIHVRT WHSDAADIDT A..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 189):

```

1  ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
51  TTTCGCGGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
15 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGC GTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
251 CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
20 351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
401 GTGTTCAAAA ATATTTTGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
501 CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
551 TTTTCGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
25 601 CAGGCAAGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
651 CAGCCTCAA CAAAGCGGCG TTATTCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCGCGC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCTTCT
30 851 TTTGGACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTCGGC
951 ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAACCTGC AACACATCA AAACGGCTGG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
35 1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAA GCATCAAAAA ATACGCTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 190; ORF32-1):

```

1  MNTPPFVCWI FCKVIDNFGD IGVSRLARV LHRELGWQVH LWTDDVSALR
40 51  ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151 LIRERDYCEA VRFDEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
201 QAGSPMTLLL AGTQIIDSLK QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLECW QTLQQHQNGW
45 351 RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR*w

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 (SEQ ID NO: 188) shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) (SEQ ID NO: 192) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
5  orf32.pep  MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVP
      |||||  |||||
orf32a  MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVX
      10      20      30      40      50      60

      70      80
10 orf32.pep  CVHQDIHVRTWHSDAADIDTA
      |||||
orf32a  CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
      70      80      90      100     110     120

```

The complete length ORF32a nucleotide sequence (SEQ ID NO: 191) is:

```

15      1  ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
      51  TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
      101  AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
      151  GCGCTTTGCC CTGATTTGCC CGATGTTNCN TGC GTTCATC AGGATATTCA
      201  TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
20      251  NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
      301  CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
      351  CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
      401  GTGTTCNAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
      451  CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
25      501  CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
      551  TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
      601  CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNCGGGGCG ANATTATCGA
      651  CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
      701  GCGATGTTTT TCAGACGGCA TCCGTCGCC TCGTCAAAAT CCCTTTCGTG
30      751  CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
      801  CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCTTCT
      851  TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
      901  GCCTTTTGGG ATAAGGACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
      951  ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
35      1001  CACAACGCCT CGAATGTTGG CAAATCCTGC AACAAATCA AAACGGCTGG
      1051  CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
      1101  ATCCGAAAAA CTCGCCCTCT TTGTTTCAAA GCATCAAAAA ATACGCTAG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 192):

```

40      1  MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
      51  ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL
      101  HIIRRHKPLW LXWEYLSAEX SNERLHXMPs PQESVKKXFW FMGFSEXSGG
      151  LIRERDYCEA VRFDGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR
      201  QAGSPLTLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
45      251  PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
      301  AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLECW QILQQHQNGW
      351  RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*

```

ORF32a (SEQ ID NO: 192) and ORF32-1 (SEQ ID NO: 190) show 93.2% identity in 382 aa overlap:

		10	20	30	40	50	60
	orf32-1.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVP					
5	orf32a	MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDVX					
		10	20	30	40	50	60
	orf32-1.pep	70	80	90	100	110	120
	orf32a	CVHQDIHVRTWHSDAADIDTAPVDPVVIETFACDLPENVLHI IRRHKPLWLNWEYLSAEE					
10	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHI IRRHKPLWLXWEYLSAEX					
		70	80	90	100	110	120
	orf32-1.pep	130	140	150	160	170	180
	orf32a	SNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS					
15	orf32a	SNERLHXMPSPQESVXKXFWFMGFSEKSGGLIRERDYCEAVRFDGALRKRMLPEKNXP					
		130	140	150	160	170	180
	orf32-1.pep	190	200	210	220	230	240
	orf32a	EWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDFVQTA					
20	orf32a	EWLLFGYRSDVWAKWLEMWRQAGSPLTLLLAGAXIIDSLKQNGVIPQDALQNDGDFVQTA					
		190	200	210	220	230	240
	orf32-1.pep	250	260	270	280	290	300
	orf32a	SVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFWHIYPQDENVHLDKLH					
25	orf32a	SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFWHIYPQDENVHLDKLH					
		250	260	270	280	290	300
	orf32-1.pep	310	320	330	340	350	360
	orf32a	AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATORLECWQTLQQHQNGWRQGAEDWSRY					
30	orf32a	AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATORLECWQILQQHQNGWRQGAEDWSRY					
		310	320	330	340	350	360
	orf32-1.pep	370	380				
	orf32a	LFGQPSAPEKLAAFVSKHQKIRX					
35	orf32a	LFGQPSASEKLAAFVSKHQKIRX					
		370	380				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 (SEQ ID NO: 188) shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) (SEQ ID NO: 194) from *N. gonorrhoeae*:

40	orf32.pep	MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLP	57
	orf32ng	MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLP	60
	orf32.pep	DVPCVHQDIHVRTWHSDAADIDTA	81
45	orf32ng	DVPFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNI IRRHKPLWLNWEYLS	120



An ORF32ng nucleotide sequence (SEQ ID NO: 193) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 194):

```

1  MVMNTYAFFV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDVVS
5  51  ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVPDAV IETFACDLPE
101 NVLNIIRRHK PLWLNWEYLS AESNERLHL MPSPQEGVQK YFWFMGFSEK
151 SGGLIRERDY REAVRFDTEA LRRRLVLPEK NAPEWLLFGY RGDVWAKWLD
201 MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
251 PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFFWHI YPQDENVHLD
301 KLHAFWDKAY GFYTPETASV HRLLSDDLNG GEALSATQRL ECGVL*

```

Further sequencing revealed the following DNA sequence (SEQ ID NO: 195):

```

1  ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
5  51  CAATTTCCGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGACCC
15  101 GCGAACTCGG TTGGCAGGTG CATTGTGGA CGGACGACGT GTCCGCCTTG
151 CGCGCGCTTT GTCCCGATTG GCCCGATGTT CCCTTCGTTT ATCAGGATAT
201 TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
251 TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAAATATT
351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCAGG
20  401 AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
451 GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA
501 AGCCCTGCGC CGGCGGCTGG TGCTGCCCCG AAAAAACGCC CCCGAATGGC
551 TGCTTTTTCG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
601 CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
25  651 CGACAGCCTC AAACAAAGCG GCGTTATTCG GCAAAACGCC CTGCAAAAtg
701 aaggcgGTGT CTTTCagacG gcacccgTcC gccttGTCAA AatcCCGTTC
751 GTGCcGCAAC AGGAcTTCGA CAAATTGCTG CacctgcCG ACTGCGCCGT
801 GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCCCT
851 TTTTTTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAAACTC
30  901 CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACGCATC
951 GGTGCACCGC CTCCTTTTCG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
1101 CGCATCCGAA AAACCTCGCC CCTTTGTTTC AAAGCATCAA AAAATACGCT
35  1151 AG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 196; ORF32ng-1):

```

1  MNTYAFFVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDVVSAL
40  51  RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVPDAVIE TFACDLPENV
101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYP WFMGFSEKSG
151 GLIRERDYRE AVRFDTEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDMW
201 QQAGSLMTLL LAGAQIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFFWHIYP QDENVHLDKL
301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLEC WQTLQQHQNG
45  351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

```

ORF32ng-1 (SEQ ID NO: 196) and ORF32-1 (SEQ ID NO: 190) show 93.5% identity in 383 aa overlap:

```

50      10      20      30      40      50      59
      orf32-1.pep  MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV

```

-196-

	orf32ng-1	MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV	10	20	30	40	50	60
5	orf32-1.pep	PCVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE	60	70	80	90	100	110
	orf32ng-1	PFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE	70	80	90	100	110	120
10	orf32-1.pep	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFPDTEALRERLMLPEKNA	120	130	140	150	160	170
	orf32ng-1	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYREAVRFPDTEALRRRLVLPEKNA	130	140	150	160	170	180
15	orf32-1.pep	SEWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQT	180	190	200	210	220	230
	orf32ng-1	PEWLLFGYRGDVWAKWLDMWQAGSLMTLLLAGAQIIDSLKQSGVIPQNALQNEGGVFQT	190	200	210	220	230	240
20	orf32-1.pep	ASVRLVKIPFVPQQDFDQLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKL	240	250	260	270	280	290
	orf32ng-1	ASVRLVKIPFVPQQDFDQLHLADCAVIRGEDSFVRTQLAGKPFFWHIYPQDENVHLDKL	250	260	270	280	290	300
25	orf32-1.pep	HAFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLCQWTLQQHQNGWRQGAEDWSR	300	310	320	330	340	350
	orf32ng-1	HAFWDKAYGFYTPETASVHRRLSDDLNGGEALSATQRLCQWTLQQHQNGWRQGAEDWSR	310	320	330	340	350	360
30	orf32-1.pep	YLFQPSASEKLAAPVSKHQKIRX	360	370	380			
	orf32ng-1	YLFQPSASEKLAAPVSKHQKIRX	370	380				

On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (SEQ ID NO: 190) (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These

experiments confirm that ORF32-1 (SEQ ID NO: 190) is a surface-exposed protein, and that it is a useful immunogen.

#### Example 24

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 197):

```

5      1  ..TTGTTCTCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
      51  GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
     101  ACGAGTGGCG GCA .ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
     151  AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTGCG TATTGTTGCT
     201  GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
    10  251  ATGCCGCTTC GGTACGCGCG GTGGAATGT TGGCATGGCT GCCGTCGAAA
     301  CTCGGTTTCC CTGTCCCCGA TCGCGGTCG GTCATCGAAG GCCGTCTGAA
     351  CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
     401  TCGCCTGCTA NGGCATCCTG CCGCGCTG..

```

15 This corresponds to the amino acid sequence (SEQ ID NO: 198; ORF33):

```

      1  ..LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
     51  SLWLCTLLGM LSVLLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
    101  LGFPVPDARS VIEGRLLNGNI ADARAWSGLL VXSIACTXGIL PRL..

```

20 Further work revealed the complete nucleotide sequence (SEQ ID NO: 199):

```

      1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
     51  AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
    101  GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GCGCGAGATG
    151  ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCGTGCGGG
    201  GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCAATTT TTTACCGGTT
    25  251  TTTCACTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
     301  GTTTTGGCGG GCGTGTGTTGG CATGAATACG CTGATGCTGG CAGTATGGTT
     351  GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
     401  CGACGTGGTT TCGGGGCAAA GACCTGTAA ATCAGGCGGT GTTGCGGCTG
    30  451  TATGCGGACG AGTGGCGGCA ACCTTCGTA CGTTGGAAAA TAGGCGCAAC
     501  GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTGCGTAT
     551  TGTTGCTGCT TTTGGTGCAG CAATATACGT TCAACTGGGA AAGCACGCTG
     601  TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
     651  GTCGAAATC GGTTCCTG TCCCCGATGC GCGGGCGGTC ATCGAAGGCC
    35  701  GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC
     751  GGCAGTATCG CCTGCTACGG CATCTGCCG CGCTGCTGG CTTGGGTAGT
     801  GTGTAAATC CTTTTGAAAA CAAGCGAAAA CGGATTGGAT TTGGAATAAGC
     851  CCTATTATCA GCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
     901  GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAAA TCATCTTGAA
    40  951  CGATGCGCGG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGCGG
    1001  AATGGTTCTGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
    1051  ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
    1101  GCGCAACTG CTTATCGGCG TCGCGGCCCA AACTGTGCCG GACCGCGGCG
    1151  TGTTCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG
    45  1201  GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
     1251  GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
     1301  CTGACAGGGC GCGCAGGAA GGGCGTTTGA AAGACCAATA A

```

This corresponds to the amino acid sequence (SEQ ID NO: 200; ORF33-1):

```

1  MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
51 IDNRNMLRET LERVLAGSFW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
101 VLAGVLGMNT LMLAVWLAML FLRVKVGRRFF SSPATWFRGK DPVNQAVLRL
151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVLR QYTFNWESTL
201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV
251 GSIACYGILP RLLAWVVCKI LLKTSENGLD LEKPYQAVI RRWQNKITDA
301 DTRRETSAV SPKIIILNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 TNREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401 VQLLAEQGLS DDLSEKLEHW RNALAECEGAA WLEPDRAAQE GRLKDQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 (SEQ ID NO: 198) shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) (SEQ ID NO: 202) from strain A of *N. meningitidis*:

```

15 orf33.pep                                     10      20      30
                                         LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR
orf33a  LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVGRRFFSSPATWFRGKDPVNQAVLR
          90      100      110      120      130      140

20 orf33.pep          40      50      60      70      80      90
                                         LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA
orf33a  LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLGDSSSVRL
          150      160      170      180      190      200

25 orf33.pep          100      110      120      130      140
                                         VEMLAWLPSKLGFPVPDARSVIEGRNLGNIA DARAWSGLLVXSIA CXGILPRL
orf33a  VEMLAWLPAKLGFPVPDARAVIEGRNLGNIA DARAWSGLLVGSIACYGILPRLLLAWAVCK
          210      220      230      240      250      260

30 orf33a  ILXXTSENGLDLEKXXXXXIRRWQNKITDADTRRETSAVSPKIVLNDAPKWAVMLETE
          270      280      290      300      310      320

```

The complete length ORF33a nucleotide sequence (SEQ ID NO: 201) is:

```

35 1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGAAGA
51 51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
101 101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGAAGATG
151 151 ATCGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGGTGCGGG
201 201 GTCGTTCTGG TGTGGGTGG GCGCGGCGAC GTTTGCGTTT NTTACCGNTT
251 251 TTTCAGTTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
40 301 GTTTTGGCGG GCGTGNTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
351 351 GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
401 401 CGACGTGGTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTTGCGGCTG
451 451 TATGCGGACG AGTGGCGGCN ACCTTCGGTA CGTTGGA AAA TAGGCGCAAC
501 501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 551 TGTGCTGCT TTTGGTGGG CAATATACGT TCAACTGGGA AAGCACGCTG
601 601 TGTGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
651 651 TGCGAAACTG GGTTTTCCCG TGCCTGATGC GCGGGCGGTC ATCGAAGGTC
701 701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTGCTGCTGCT GCTGCTGCTG

```

5  
10

```

751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGCGGT
801 ATGCAAAATC CTTNTGNAAA CAAGCGAAAA CGGCTTGGAT TTGGAAAAGC
851 NNNNNNTCN NNCGNTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA
951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG
1001 AATGGTTCTGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGGCG TCGCGCCCA AACTGTGCCC GACCGCGCGC
1151 TGTTGCGGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
1301 CCGACAGAGC GCGCAGGAA GGCCGTCTGA AAACCAACGA CCGCACTTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 202):

15  
20

```

1  MLNPSRKLVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
51  IDRNRMLRET LERVAGSEFW LWVAAATFAF XTXFSVTYLL MDNQGLNFFL
101 VLAVVGMNT LMLAVWLAML FLRVKVGRRF SSPATWFRGK DPNQAVLRL
151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201 LGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301 DTRRETSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKQVA
351 ANREQVALE TELKQKPAQL LIGVRAQTPV DRGVLRIQVR LSEAAQGGAV
401 VXLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRLKTNDRT*

```

25 ORF33a (SEQ ID NO: 202) and ORF33-1 (SEQ ID NO: 200) show 94.1% identity in 444 aa overlap:

30  
35  
40  
45  
50

```

              10      20      30      40      50      60
orf33a.pep    MLNPSRKLVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDRNRMLRET
              |||
orf33-1        MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDRNRMLRET
              10      20      30      40      50      60

              70      80      90      100     110     120
orf33a.pep    LERVAGSEFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
              |||
orf33-1        LERVAGSEFWLWVVAATFAFFTGFSTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
              70      80      90      100     110     120

              130     140     150     160     170     180
orf33a.pep    FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
              |||
orf33-1        FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
              130     140     150     160     170     180

              190     200     210     220     230     240
orf33a.pep    VSVLLLLLVRQYTFNWESTLLGDSSSVRLVEMLAWLPAKLGFPVPDARAVIEGRNLGNIA
              |||
orf33-1        VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
              190     200     210     220     230     240

              250     260     270     280     290     300
orf33a.pep    DARAWSGLLVGSIACYGILPRLLAWAVCKILXXTSENGLDLEKXXXXXXIIRRWQNKITDA
              |||
orf33-1        DARAWSGLLVGSIACYGILPRLLAWVCKILLKTSENGLDLEKPYQAVIRRWQNKITDA
              250     260     270     280     290     300

```

-200-

		310	320	330	340	350	360
	orf33a.pep	DTRRET	VS	SAVSPKIVLNDAPKWAVMLETEWQDGEWFEGR	LAQEWLDKGVAANREQVA	AALE	
5	orf33-1	DTRRET	VS	SAVSPKII	LNDAPKWAVMLETEWQDGEWFEGR	LAQEWLDKGVA	TNREQVA
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf33a.pep	TELKQKPAQLLIGVRAQTV	PDRGVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW				
10	orf33-1	TELKQKPAQLLIGVRAQTV	PDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW				
		370	380	390	400	410	420
		430	440	450			
	orf33a.pep	RNALTECGAAWLEPDRAAQEGRLKTNDR	TX				
15	orf33-1	RNALAECGAAWLEPDRAAQEGRLKDQX					
		430	440				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 (SEQ ID NO: 198) shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) (SEQ ID NO: 204) from *N. gonorrhoeae*:

20	orf33.pep		LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFFLVLAGVLMNTLM	LAVWLATLFLRVKVGRRFFSSPATWFRGKGPVNQAVLR	100
	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLV	RQYTFNWESTLLSNAASVRA	90
	orf33ng	LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLV	RQYTFNWESTLLSNAASVRA	160
25	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRLNGNIADARAWSGLLVXS	IACXGILPRL	143
	orf33ng	VEMLAWLPSKLGFPVPDARAVIEGRLNGNIADARAWSGLLVGS	IVCYGILPRL	220

30 An ORF33ng nucleotide sequence (SEQ ID NO: 203) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 204):

	1	MIDRDRMLRD	TLERV	RAGSF	WLWVV	VASMM	FTAGF	SGTYL	LMDNQGLNFF
	51	LVL	AGVLMN	TLML	AVLAT	LFLRV	KVGRF	FSSPATWFRG	KGPVNQAVLR
	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSVLLLLLV	RQYTFNWEST			
35	151	LLSNAASVRA	VEMLAWLPSK	LGFPVPDARA	VIEGRLNGNI	ADARAWSGLL			
	201	VGSIVCYGIL	PRL	LAWVVK	ILLKTSENGL	DLEKTYQAV	IRRWQNKITD		
	251	ADTRRET	VSA	VSPKIVL	NDA	PKWALMLETE	WQDQWFEGR	LAQEWLDKGV	
	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGA			
	351	VVQLLAEQGL	SDDLSEKLEH	WRNALTECGA	AWLEPDRAV	Q	EGRLKDQ*		

40 Further sequence analysis revealed the following DNA sequence (SEQ ID NO: 205):

	1	ATGTTGaatC	CATCCCgaAA	ACTGgttgag	ctGgTCCgtA	Ttttgaataa
	51	aggggggtTTT	attttcagcg	gcgatacctgt	gcaggcgacg	gaggctttgc
	101	gccgcgtgga	cggcAGTACG	GAGgAaaaaa	tcttcgctcg	GGCGGAGAtg
	151	atcgACAGGg	accgtatgtt	gcgggACaCg	TtggaaacGTG	TGCGTGCGgg

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65  
 70  
 75  
 80  
 85  
 90  
 95  
 100  
 105  
 110  
 115  
 120  
 125  
 130

201 gtcgtTctgG TTATGGGTGG TggtggCatC gATGATGtTt aCCGCCGGAT  
 251 TTTCAGgcac ttatCttCTG ATGGACaatC AGGGGcTGAA TtTCTTTTAA  
 301 GTTTTggcgG GAGTGTtggG CATGaatacG ctgATGCTGG CAGTATGGtt  
 351 gGCAACGTTG TTCCTGCGCG TGAAAGTGGG ACGGTTTTTC AGCAGTCCGG  
 401 CGACGTGGTT TCGGGGCAAA GGCCCTGTAA ATCAGGCGGT GTTGC GGCTG  
 451 TATGCGGACC AGTGGCGGCA ACCTTCGGTA CGATGGAAAA TAGGCGCAAC  
 501 GGCGCACAGC TTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCCGTAT  
 551 TGCTGCTGCT TTTGGTGC GG CAATATACGT TCAACTGGGA AAGCACGCTG  
 601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTGG CATGGCTGCC  
 651 GTCGAAACTC GGTTTCCCTG TCCCCGATGC GCGGGCGGTC ATCGAAGGTC  
 701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC  
 751 GGCAGTATCG TCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGTAGT  
 801 GTGTAAATC CTTTTGAAAA CAAGCGAAAA CGGattgGAT TTGGAAAAAA  
 851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG  
 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCcgaAAA TCGTCTTGAA  
 951 CGATGCGCGG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC  
 1001 AATGGTTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC  
 1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC  
 1101 GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG  
 1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG  
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT  
 1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC  
 1301 CTGACAGGGT GGCGCAGGAA GGCCGTTTGA AAGACCAATA A

25 This encodes a protein having amino acid sequence (SEQ ID NO: 206; ORF33ng-1):

1 MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRRAEM  
 51 IDRDRMLRDT LERV RAGSFW LWVVVAMMF TAGFSGTYLL MDNQGLNFFL  
 101 VLAGVLGMNT LMLAVWLATL FLRVKVRGFF SSPATWFRGK GPNQAVLRL  
 151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL  
 201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV  
 251 GSIVCYGILP RLLAWVVCKI LLKTSENGLD LEKTYQAVI RRWQNKITDA  
 301 DTRRETSAV SPKIVLNDAP KWALMLETEW QDQWFEGRL AQEWLDKQVA  
 351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV  
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ\*

ORF33ng-1 (SEQ ID NO: 206) and ORF33-1 (SEQ ID NO: 200) show 94.6% identity in 446 aa overlap:

40  
 45  
 50

orf33-1.pep	MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRMLRET
orf33ng-1	MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGSTEEKIFRRAEMIDRDRMLRDT
orf33-1.pep	LERVRAGSFWLWVVAATFAFFTGFSTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLAML
orf33ng-1	LERVRAGSFWLWVVVAMMFTAGFSGTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLATL
orf33-1.pep	FLRVKVRGFFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
orf33ng-1	FLRVKVRGFFSSPATWFRGKGPVNQAVLRLYADQWRQPSVRWKIGATAHSLWLCTLLGML

		190	200	210	220	230	240
	orf33-1.pep	VSVLLLLLV	RQYTFN	WESTLLS	NAASVRA	VELAWLPS	KLGFVPD
	orf33ng-1	VSVLLLLLV	RQYTFN	WESTLLS	NAASVRA	VELAWLPS	KLGFVPD
5		190	200	210	220	230	240
	orf33-1.pep	DARAWSG	LLVGSIA	CYGILPR	LLAWVV	CKILLKT	SENGLD
	orf33ng-1	DARAWSG	LLVGSIA	CYGILPR	LLAWVV	CKILLKT	SENGLD
10		250	260	270	280	290	300
	orf33-1.pep	DARAWSG	LLVGSIA	CYGILPR	LLAWVV	CKILLKT	SENGLD
	orf33ng-1	DARAWSG	LLVGSIA	CYGILPR	LLAWVV	CKILLKT	SENGLD
		250	260	270	280	290	300
	orf33-1.pep	DTRRET	VSAVSP	KIILND	APKWAV	MLETWQ	DGEWFE
	orf33ng-1	DTRRET	VSAVSP	KIILND	APKWAV	MLETWQ	DGEWFE
15		310	320	330	340	350	360
	orf33-1.pep	DTRRET	VSAVSP	KIILND	APKWAV	MLETWQ	DGEWFE
	orf33ng-1	DTRRET	VSAVSP	KIILND	APKWAV	MLETWQ	DGEWFE
		310	320	330	340	350	360
	orf33-1.pep	TELKQK	PAQLLI	GVRAQT	VPDRGV	LRQIVRL	SEAAQGG
	orf33ng-1	TELKQK	PAQLLI	GVRAQT	VPDRGV	LRQIVRL	SEAAQGG
20		370	380	390	400	410	420
	orf33-1.pep	TELKQK	PAQLLI	GVRAQT	VPDRGV	LRQIVRL	SEAAQGG
	orf33ng-1	TELKQK	PAQLLI	GVRAQT	VPDRGV	LRQIVRL	SEAAQGG
		370	380	390	400	410	420
	orf33-1.pep	RNALAEC	GAAWLE	PDRAAQ	EGRLKD	QX	
	orf33ng-1	RNALAEC	GAAWLE	PDRAAQ	EGRLKD	QX	
25		430	440				

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 25

30 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 207):

	1	..CAGAAGAGTT	TGTCGAGAAT	TTCTTTATGG	GGTTTGGGCG	GCGTGTTTTT
	51	CGGGGTGTCC	GGTCTGGTAT	GGTTTCTTTT	GGGCGTTTCT	TT.GAGTGC
	101	CCTGTTTTTC	GGGTGTTTCT	TTTCGGGGTT	CGGGACGGGG	GACGTTTGTG
	151	GGCAGTACGG	GGGTTTCTTT	GAGTGTGTTT	TCAGCTTGTG	TTCC.GGCGT
35	201	CGTCCGGCTG	CCTGTCGGTT	TGAGCTGTGT	CGGCAGGTTG	CG..GTTTGA
	251	CCCGGTTTTT	CTTGGGTGCG	GCAGGGGACG	TCATTCTCCT	GCCGCTTTTCG
	301	TCTGTGCCGT	CCGGCTGTGC	GGGTTCCGAT	GAGGCGGCGT	GGTGGTGTTC
	351	GGGTTGGGCG	GCATCTTGTT	CCGACTACGC	CGTTTGGCAG	CCAGAATTTCG
	401	GTTTCGCGGG	GGCTGTCCGT	GTGTTGCGGT	TCGGCTTGAA	GGGTTTTGTG
40	451	GTCC..				

This corresponds to the amino acid sequence (SEQ ID NO: 208; ORF34):

	1	..QKSLSRISLW	GLGGVFFGVS	GLVWFSLGVS	XECACFSGVS	FRGSGRGTFV
	51	GSTGVSLSVF	SACVXGVVRL	PVGLSCVGRL	XXLTRFFLGA	AGDVILLPLS



101 SVPSGCAGSD EAAWCSGWA ASCPTTPFGS QNSVSRGLSV CCGSA\*RVLS  
 151 S..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 209):

```

5      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGcKGGTG TGCCTGCCGT
      51  GCCGGGTCAG AATAGGTTGT CCAGAATTTC TTTATGGGGT TTGGGCGGGC
     101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
     151  GGCTGCGCCT GTTTTTTCGGG TGTTTCTTTT CGGGGTTTCGG GACGGGGGAC
     201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTCA GCTTGTGTTT
    10  251  CGGCGTCGTC CGGCTGCCTG TCGGTTTGAG CTGTGTCGGC AGGTTGCGGT
      301  TTGACCCGGT TTTTCTTGGG TCGGCGAGG GACGGCAGTC CGCTGCCGCT
      351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGTGTTG
      401  GTTCGGGTTG GCGGCGATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAA
      451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
    15  501  GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
      551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
      601  AGCCTGAAGG GTTTGTTTCGG TTTTTTTGCC ATTTTGATTG TGCTTTTGGG
      651  GTGTCGGGCA ATGCCGTCTG AAGGCGGTTT AGACGGCATT GCCGAGTCAG
      701  CGTTGGACGT AGTTTTGTTA GAGGGTGATG ACTTTTTGTA CGCCGACGGT
    20  751  GGTGCTGACT TTTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
      801  CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
      851  GTGGCGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCA
      901  AGTGTCGCCG GCGATGTCGC CGGCAGTGC GCGCAGGGAG GCGACGGTAA
      951  TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
    25 1001  ACGAACTGTT TTTCGCCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
      1051  GCGGTTGTAG CCGACGACGG AGATTGCGG CGTGTAGCCT TTGGTTTGGT
      1101  TGTTTTGGCG CAGATAGGAA CGGGCGGTGG TTTTCGATAC CAACGCCATA
      1151  ACGTTGTCGT CGGTTTGCGC GCCCGTGGTT CGGCGGTCCA CGGCGGATTT
      1201  CGGCGCCGAC GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
    30 1251  CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT
      1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
      1351  CATGCCGTCT GA
  
```

This corresponds to the amino acid sequence (SEQ ID NO: 210; ORF34-1):

```

35      1  MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLGVSL
      51  GCACFSGVSF RSGRGTFFVG STGVSLSVFS ACVPASSGCL SV*AVSAGCG
     101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTPFGSQN
     151  SVSRGLSVCC GSA*RVLSPF GLNVLTMPIA NAPMAAIQMS NTARIRSLGV
     201  SLKGLFGFFA ILIVLLGCRA MPSEGGSDGI AESALDVVLV EGDDFLYADG
    40  251  GADFLGNLRL FFGGEDAHNV GYVAVGNDFD ARLCGGADAQ QRGADFGCVP
      301  SVAGDVAGSA RQGGDGNIVV HAFGGLFGTC NLTDELFFAF GGDLEQQQV
      351  AVVADDGDLG RVAFGLVLA QIGTGGGFD TQRHNVVGLR AGGSAVDGGF
      401  RADGGASDYC ADAAAKGKAE NGGNQGADGV RFGFHRVLPF LGVSDGIALR
      451  HAV*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF34 (SEQ ID NO: 208) shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) (SEQ ID NO: 212) from strain A of *N. meningitidis*:

```

5      orf34.pep      10      20      30      40      50      60
                QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE-----CAC
                |||||
      orf34a      MMXPXIMLPWIIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWFSLGVSXSLGVSXGXCAC
                10      20      30      40      50      60

      40      50      60      70      80      90
      orf34.pep      FSGVSFRGSGRGTFFVGSTGVSLSVFSACVXGVVRLPVGLSCVGRLLX-----LTRFFFLGA
                |||||
      orf34a      FSGVSFRGSGRGTFFVGSTGVSLSVFSACA-----PASSGCLSVXAVSAGCGLTRXFXFGA
                70      80      90      100      110

      100      110      120      130      140      150
      orf34.pep      AGDVILLPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLS
                |||||
      orf34a      AGDGSPLPLSSVPSGCAGADEEAXXCSGWAASCPTTPFGSQNSVSRGLSVCCGSVVRVLS
                120      130      140      150      160      170

      orf34.pep      S

      orf34a      PFGXNVLTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSD
                180      190      200      210      220      230

```

The complete length ORF34a nucleotide sequence (SEQ ID NO: 211) is:

	1	ATGATGATNC	CGTTNATAAT	GCTTCTTGG	ATTGCGGGTG	TCCTGCCGCT
	51	GCCGGGTCAG	AAGAGGTTGT	CGAGAANTTC	TTTATGGGGT	TTAGCGGCN
25	101	TGTTTTTCGG	GGTGTCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTNTT
	151	TCTTTGGGTG	TTTCTNTGGG	CTGTGCCTGT	TTTTCGGGTG	TTTCTTTTTCG
	201	GGGTTTCGGA	CGGGGGACGT	TTGTGGGCAG	TACNGGGGTT	TCTTTGAAGT
	251	TGTTTTTCAG	TTGTGCTCCG	GCGTCGTCCG	GCGTCTCTGC	GGTTTNAAGT
	301	GTGTCGCGAG	GTTGCGGTTT	GACCCGNTT	TTCTTNGGTG	CGGCAGGGGA
30	351	CGGCAGTCCG	CTGCCGCTTT	CGTCTGTGCC	GTCCGGCTGT	GCGGGTGCGG
	401	ATGAGGAGGC	GTNGTNGTGT	TCGGGTGGG	CGGCATCTTG	TCCGACTACG
	451	CCGTTTGCA	GCCAGAATTC	GGTTTCGCGG	GGGCTGTGCG	TGTGTTGCGG
	501	TTCGGNTTGG	AGGGTTTTGT	CNCCGTTTCG	GTNGAATGTG	CTGACGATGC
	551	CTATTGCCAA	TGCCCCGATG	GCGGTGATAC	AGATGAGCAA	TACGGCCGCT
35	601	ATCAGGAGTT	TGGGGTCAG	CCTGAAGGGT	TTGTTCTNGT	TTTTTGCCAT
	651	TTTGATTGTG	CTTTTGGGGT	GTCGGGCAAT	GCCGTCTGAA	GGCGGTTTCA
	701	ACGGCATTGC	CGAGTCAGCG	TTGGACGTAG	TTTNGGTAGA	GGGTGATGAC
	751	TTTTTGTACG	CCGACGGTGG	TGCTGACTTT	TTGGGTAATC	TGCGCTGTG
	801	CTTCGGGGGT	GAGGATGCCC	ATAACGTAGG	TTACGTTGCC	GTAGGTAACG
40	851	ATTTTGACGC	GCGCCTGTGT	GGCGGGGCTG	ATGCCCAACA	GCGTGGCGCG
	901	GACTTTGGAT	GTGTTCCAAG	TGTCGCGCGG	GATGTCGCGG	CGAGTGGCGG
	951	GCAGGGAGGC	GACGGTAATG	TANTTTGTACA	CGCCTTTCGG	GGCCTGTTCG
	1001	GAACGTGCAA	TCTGACCGAC	GAACGTGTTT	TCGCTTTCGG	TGGCGACTTG
	1051	TCCGAGCAGC	AGCAGGTGGC	GGTTGTAGCC	GACAACGGAG	ATTTGGGGCG
45	1101	TGTANCTTT	GGTTTGTTG	TTTTGGCGCA	GATAGGAGCG	GGCGGTGGTT
	1151	TCGATACGCA	GCGCCATTAC	GTTGTCGTG	GTTCGCGCGC	CGGTGGTTTCG
	1201	GCGGTCGACG	GCGGATTTTCG	CGCCGACCGC	CGCGCCGCGG	ACGACTGCGC
	1251	TGACGTCAGC	GCCGAGGGCA	AGGCTGAGGA	CGGCGGCAGT	CAGGGTGCGG
	1301	ACGGGTGTCG	GTTTGGGTTT	CATCGGTGTC	TTCTTTCTT	GGGCGTTTCA
	1351	GACGGCATTG	CTTTGCGCCA	TGCCGTCTGA		

This encodes a protein having amino acid sequence (SEQ ID NO: 212):

1 MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLGVSX  
51 SLGVSXGCAC FSGVSFRGSG RGTFTVGSTGV SLSVFSACAP ASSGCLSVXA

101 VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT  
 151 PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR  
 201 IRS LGVSLKG LFXFFAILIV LLGCRAMPSE GGSDGIAESA LDVVXVEGDD  
 251 FLYADGGADF LGNLR LFFGG EDHNVGYVA VGNDFDARLC GGADAQQRGA  
 301 DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL  
 351 SEQQQVAVVA DNGDLGRVXF GLVVL AQIGA GGGFDTQRHY VVVGXRAGGS  
 401 AVDGGFRADR RAADD CADAA AEGKAEDGGS QGADGVRFGF HRVLPFLGVS  
 451 DGIALRHAV\*

10 ORF34a (SEQ ID NO: 212) and ORF34-1 (SEQ ID NO: 210) show 91.3% identity in 459 aa overlap:

		10	20	30	40	50	60
	orf34a.pep	MMXPXIMLPW	IAGVPAVPGQ	KRLSRXSLW	GLGGXFFGV	SGLVWFS	LGVSXSLGVSXGCAC
15	orf34-1	MMMPPFIMLPW	IAGVPAVPGQ	NRLSRISLW	GLGGVFFGV	SGLVWFS	LGVS-----GCAC
		10	20	30	40	50	
	orf34a.pep	70	80	90	100	110	120
	orf34a.pep	FSGVSFRGSGRGT	TFVGSTGVSL	SVFSACAPASS	GCCLSVXAVS	AGCGLTRXFX	GGAAGDGSP
20	orf34-1	FSGVSFRGSGRGT	TFVGSTGVSL	SVFSACVPASS	GCCLSVXAVS	AGCGLTRFFL	GGAAGDGSP
		60	70	80	90	100	110
	orf34a.pep	130	140	150	160	170	180
	orf34a.pep	LPLSSVPSGCAG	ADEEAXXC	SGWAASCPTT	PFGSQNSVSR	GLSVCCGSVW	RVLSPFGXNV
25	orf34-1	LPLSSVPSGCAG	SDEAAWCS	GWAAASCPTT	PFGSQNSVSR	GLSVCCGSAX	RVLSPPGLNV
		120	130	140	150	160	170
	orf34a.pep	190	200	210	220	230	240
	orf34a.pep	LTMPIANAPMA	VIQMSNTAR	IRSLGVSLK	GLFXFFAIL	IVLLGCRAMP	SEGGSDGIAESA
30	orf34-1	LTMPIANAPMA	AIQMSNTAR	IRSLGVSLK	GLFGFFAIL	IVLLGCRAMP	SEGGSDGIAESA
		180	190	200	210	220	230
	orf34a.pep	250	260	270	280	290	300
	orf34a.pep	LDVVXVEGDD	FLYADGGAD	FLGNLRLFF	GGEDAHNV	GYVAVGN	DFDARLCGGADAQQRGA
35	orf34-1	LDVVLVEGDD	FLYADGGAD	FLGNLRLFF	GGEDAHNV	GYVAVGN	DFDARLCGGADAQQRGA
		240	250	260	270	280	290
	orf34a.pep	310	320	330	340	350	360
	orf34a.pep	DFGCVPSVAG	DVAGSARQ	GGDGNVXV	HAFGGLFG	TCNLTD	ELFLAFGGDLSEQQQVAVVA
40	orf34-1	DFGCVPSVAG	DVAGSARQ	GGDGNIVV	HAFGGLFG	TCNLTD	ELFFAFGGDLSEQQQVAVVA
		300	310	320	330	340	350
	orf34a.pep	370	380	390	400	410	420
	orf34a.pep	DNGDLGRVXF	GLVVL AQ	IGAGGGFDT	QRHYVVVG	XRAGGS	AVDGGFRADRRRAADD CADAA
45	orf34-1	DDGDLGRVAF	GLVVL AQ	IGTGGGFDT	QRHNVVVG	LRAGGS	AVDGGFRADGGASDYCADAA
		360	370	380	390	400	410
	orf34a.pep	430	440	450	460		
	orf34a.pep	AEGKAEDGGS	QGADGVR	FGFHRVLP	FLGVSD	DGIALRHAVX	
50	orf34-1	AKGKAENGGN	QGADGVR	FGFHRVLP	FLGVSD	DGIALRHAVX	

420 430 440 450

Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 (SEQ ID NO: 208) shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) (SEQ ID NO: 214) from *N. gonorrhoeae*:

5	orf34.pep	QKSLSRISLWGLGGVFFGVSGLVWFSLVGSXE-----CAC	35
	orf34ng	MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC	60
	orf34.pep	FSGVSFRGSGRGTFVVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFLGA	90
10	orf34ng	FSGVSFRGSGWGAFAVGSTGVSLSVFSACVP----VPVNESAARAASEGR--GLTRFFLGA	114
	orf34.pep	AGDVILLPLSSVPSGCAGSDEAAWWC SGWAASCTTPFGSQNSVSRGLSVCCGSAXRVLS	150
	orf34ng	AGDGSPLPLSSVPSGCAGSDEAAWWC SGWAASCTAPFGSQNSVSRGLSVCCGSVWRVLS	174
	orf34.pep	S	175
15	orf34ng	PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD	234

The complete length ORF34ng nucleotide sequence (SEQ ID NO: 213) is:

	1	ATGATGATGC	CGTTCATAAT	GCTTCCTTGG	ATTGCGGGTG	TGCCTGCCGT
	51	GCCGGGTCAA	AAGAGGTTGT	CGAGAATCTC	TTTATGGGGT	TTGGCCGGCG
20	101	TGTTTTTCGG	GGTGTCCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTTTT
	151	TCTTTGGGTG	TTTCTTTGGG	CTGCGCCTGT	TTTTCGGGTG	TTTCTTTTCG
	201	GGGTTTCGGA	TGGGGGGCGT	TTGTGGGCAG	TACGGGGGT	TCTTTGAGTG
	251	TGTTTTCAGC	TTGTGTTCCG	GTGCCGGTTA	ACGAATCGGC	TGCCCCGGCC
	301	GCATCCGAAG	GGCGCGGTTT	gACCCGGTTT	TTCTTGGGTG	CGGCAGGGGA
25	351	CGGCAGTCCG	CTGCCGCTTT	CTTCTGTGCC	GTCCGGCTGT	GCGGGTTCCG
	401	ATGAGGCGGC	GTGGTGGTGT	TCGGGTTGGG	CGGCATCTTG	TCCGACGGCG
	451	CCGTTTGGCA	GCCAGAATTC	GGTTTCGCGG	GGGCTGTCGG	TGTGTTGCGG
	501	TTCGGTTTGG	AGGGTTTTGT	CGCCGTTCCG	GTTGAATGTG	CTGACGATGC
	551	CTACTGCCAA	TGCGCCGATG	GCGGTGATAC	AGATGAGCAA	TACGGCGCGT
30	601	ATCAGGAGTT	TGGGGGTCAG	CCTGAAGGGT	TTGTTCCGGT	TTTTTGCCAT
	651	TTTGATTGTG	CTTTTGGGGT	GTCGGGCAAT	GCCGTCTGAA	GGCGTTTCAG
	701	ACGGCATTGC	CGAGTCAGCG	TTGGACGTAG	TTTTGGTAGA	GGGTAATGAC
	751	TTTTTGACG	CCGAcggtGG	TGCTGACTTT	TTGGGTAATC	TGCGCCTGTT
	801	CTTCGGGGGT	GAGGATGCCC	ATAACGTAGG	TTACATTGCC	GTAGGTAATG
35	851	ATTTTGACGC	GCGCCTGTGT	AGCGGGGCTG	ATGCCCAGCA	GcgtgGCGCG
	901	GACTTTGGAC	GTGTTCCAAG	TGTCGCCGGC	GATGTCGCCC	GCAGTGCGCG
	951	GCAGGGAGGC	GACGGTAATG	TAGTTGTATA	CGCCTTCGGC	GGCCTGTTTCG
	1001	GAACGTGCAA	TCTGACCGAC	GAAGTGTGTT	TGCGCTTCGG	TGGCGACTTG
	1051	TCCGAGCAGC	AGCAGGTGGC	GTTGTAGACC	GACGACGGAG	ATTTGGGGCG
40	1101	TGTAGCCTTT	GGTTTGGTTG	TTTTGGCGCA	GGTAGGAACG	GGCGGTGGTT
	1151	TCGATACGCA	ACGCCATAAC	GTtgtCATCG	GTTtgcgcgc	CGGTGGTTcg
	1201	gCGGTCGATG	ACGATTTTGT	CGCCGACGGC	GGCCCCGCGC	ACGACTGCGC
	1251	TGAAGCAGCC	GCCGAGGGCA	AGGCTGAGGA	CGGCGGCAAT	CAGGGTGCGG
	1301	ACGGTGTGTG	GTTTGGGTTT	CATCGGGGAC	TTCTTTCTT	GGGCGTTTCA
45	1351	GACGGCATTG	CTTTGCGCCA	TGCCGTCTGA		

This encodes a protein having amino acid sequence (SEQ ID NO: 214):

1 MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LVWFSLGVSF  
 51 SLGVSLGCAC FSGVSFRGSG WGA FVGSTGV SLSVFSACVP VPVNESAARA  
 101 ASEGRLTRF FLGAAGDGSP LPLSSVPSGC AGSDEAAWWC SGWAASCPTA  
 151 PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMP TANAPM AVIQMSNTAR  
 5 201 IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GGSDGIAESA LDVVLVEGND  
 251 FLYADGGADF LGNLR LFFGG EDAHNVGYIA VGND F DARLC SGADAQQRGA  
 301 DFGRVPSVAG DVAR SARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL  
 351 SEQQQVAVVA DDGDLGRVAF GLVVL AQVGT GGGFD TQRHN VVIGLRAGGS  
 401 AVDDGFCADG GPADDCAEAA AEGKAEDGDN QGADGVWFGF HRGLPFLGV S  
 10 451 DGIALRHAV\*

ORF34ng (SEQ ID NO: 214) and ORF34-1 (SEQ ID NO: 210) show 90.0% identity in 459 aa overlap:

15	orf34-1.pep	10 20 30 40 4 50	MMMPFIMLPW IAGVPAVPGQ NRLSRISLWGLGGVFFGVSGLVWFSLGV S-----LGCAC
	orf34ng	10 20 30 40 50 60	MMMPFIMLPW IAGVPAVPGQ KRLSRISLWGLAGVFFGVSGLVWFSLGV SFSLSGVSLGCAC
20	orf34-1.pep	60 70 80 90 100 110	FSGVSFRGSGRGT FVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
	orf34ng	70 80 90 100 110 120	FSGVSFRGSGWGA FVGSTGVSLSVFSACVP VPVNESAARAASEGRLTRFFLGAAGDGSP
25	orf34-1.pep	120 130 140 150 160 170	LPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV
	orf34ng	130 140 150 160 170 180	LPLSSVPSGCAGSDEAAWWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV
30	orf34-1.pep	180 190 200 210 220 230	LTMPIANAPMAA IQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA
	orf34ng	190 200 210 220 230 240	LTMP TANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA
35	orf34-1.pep	240 250 260 270 280 290	LDVVLVEGDDFLYADGGADFLGNLR LFFGGEDAHNVGYVAVGND F DARLCGGADAQQRGA
	orf34ng	250 260 270 280 290 300	LDVVLVEGNDFLYADGGADFLGNLR LFFGGEDAHNVGYIAVGND F DARLCGADAQQRGA
40	orf34-1.pep	300 310 320 330 340 350	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTD ELFFAFGGDLSEQQQVAVVA
	orf34ng	310 320 330 340 350 360	DFGRVPSVAGDVAR SARQGGDGNVVVYAFGGLFGTCNLTD ELFFAFGGDLSEQQQVAVVA
45	orf34-1.pep	360 370 380 390 400 410	DDGDLGRVAFGLVVL AQIGTGGGFD TQRHNVVGLRAGGSAVDGGFRADGGASDYCADAA
	orf34ng	370 380 390 400 410 420	DDGDLGRVAFGLVVL AQVGTGGGFD TQRHNVVIGLRAGGSAVDDGFCADGGPADDCAEAA
50	orf34-1.pep	420 430 440 450	AKGKAENGNGQADGVRFGRVLPFLGVSDGIALRHAVX

orf34ng  
 AEGKAEDGGNQGADGVWFGFHRGLPFLGVSDGIALRHAVX  
 430 440 450 460

- 5 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 26

- 10 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 215):

1 ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT  
 51 CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG  
 101 CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC  
 15 151 CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAAATCCAA GCCGAGCTGG  
 201 AGAAAAAAGG CTACACCGTC AAATGGTTCG AGTTTACCGA CTATGTACGC  
 251 CCGAATCTGG CATTTGGCTGA GGGCGAGTTG

This corresponds to the amino acid sequence (SEQ ID NO: 216; ORF4):

20 1 MKTFFKTLA AALALILAAC G.QKDSAPAA SASAAADNGA AKKEIVFGTT  
 51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

Further sequence analysis revealed the complete nucleotide sequence (SEQ ID NO: 217):

25 1 ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT  
 51 CGCCGCCTGC GGCGGTCAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG  
 101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC  
 151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA  
 201 GAAAAAAGGC TACACCGTCA AACTGGTTCG GTTTACCGAC TATGTACGCC  
 251 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTCACACAC  
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG  
 351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCG GGCAAGCTGA  
 30 401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC  
 451 CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTGGAT  
 501 CAAACTCAA GACGGCATCA ATCCGTTGAC CGCATCCAA GCGGACATCG  
 551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTGAAGC CGCGCAACTG  
 601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC  
 35 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT  
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA  
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA  
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG  
 851 GCGCAGCCAA ATAA

40 This corresponds to the amino acid sequence (SEQ ID NO: 218; ORF4-1):

45 1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT  
 51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH  
 101 KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND  
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL

201 PRSRADVDFV VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ  
251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK\*

Computer analysis of this amino acid sequence gave the following results:

### 5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 (SEQ ID NO: 216) shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) (SEQ ID NO: 220) from strain A of *N. meningitidis*:

		10	20	30	40	50	59
10	orf4.pep	MKTFFKTL SAAALALILAACG-QKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE					
	orf4a	MKTFFKTL SAAALALILAACGGQKDSAPAASASAAAADNGAAXKEIVFGTTVGDFGDMVKE					
		10	20	30	40	50	60
15	orf4.pep	60	70	80	90		
		QIQAELEKKGYTVKLVEFTDYVRPNLALAEGL					
	orf4a	XIQPELEKKGYTVKLVEXTDYVRXNLALAEGLDINVXQHXXYLDXKKXHNLDITXVXQ					
		70	80	90	100	110	120
20	orf4a	VPTAPLGLYPGKLKSLXXVKXGSTVSAPNDPXXFXRVLVMLDELGXIKLKDXIXXXXXXX					
		130	140	150	160	170	180

The complete length ORF4a nucleotide sequence (SEQ ID NO: 219) is:

	1	ATGAAAACCT	TCTTCAAAAC	CCTTTCCGCC	GCCGCACTCG	CGCTCATCCT
	51	CGCCGCTG	GGCGGTCAA	AAGATAGCG	GCCCGCCGA	TCCGCTTCTG
25	101	CCGCCGCCA	CAACGGCGC	GCGAANAAAG	AAATCGTCTT	CGGCACGACC
	151	GTCGGGACT	TCGGCGATAT	GGTCAAAGAA	CANATCCAAC	CCGAGCTGGA
	201	GAAAAAAGC	TACACCGTCA	AACTGGTCTGA	GTNTACCGAC	TATGTGCGCN
	251	CGAATCTGG	ATTGGCTGAG	GGCGAGTTGG	ACATCAACGT	CTTNCAACAC
	301	ANACNCTAT	TTGACGACTN	CAAAAAANAA	CACAATCTGG	ACATCACCNN
	351	AGTCTTNCA	GTGCCGACCG	CGCCTTTGGG	ACTGTACCCG	GGCAAGCTGA
30	401	AATCGCTGA	NNAAGTCAA	GANGGCAGCA	CCGTATCCGC	GCCCAACGAC
	451	CCGTNNNACT	TCGNCCGCGT	CTTGGTGATG	CTCGACGAAC	TGGGTNGAT
	501	CAAACTCAA	GACNGCATCA	NNNNGNNGNN	NNNANCNANA	NNNGANANN
	551	NNNNANNNNT	NNNNNNNNNN	NNNNNCNNCG	NNNNNNNANN	NNNNNNNNNN
	601	NCGNNTNNNN	NNGCNNNNNT	NNANNNTNNN	NNCNCNNNNN	NNNNNTNNNN
35	651	NANNANNAGC	GGCATGAAGC	TGACCGAAGC	CCTGTTCCAA	GAACCGAGCT
	701	TTGCCTATGT	CAACTGTCT	GCCGTCAAAA	CCGCCGACAA	AGACAGCCAA
	751	TGGCTTAAAG	ACGTAACCGA	GGCCTATAAC	TCCGACGCGT	TCAAAGCCTA
	801	CGCGCACAAA	CGCTTCGAGG	GCTACAAATC	CCCTGCCGCA	TGGAATGAAG
40	851	GCGCAGCCAA	ATAA			

This is predicted to encode a protein having amino acid sequence (SEQ ID NO: 220):

	1	MKTFFKTL S	AALALILAAC	GGQKDSAPAA	SASAAAADNGA	AXKEIVFGTT
	51	VGDFGDMVKE	XIQPELEKKG	YTVKLVEXTD	YVRXNLALAE	GELDINVXQH
	101	XXYLDXKKX	HNLDITXVXQ	VPTAPLGLYP	GKLKSLXXVK	XGSTVSAPND
45	151	PXXFXRVLVM	LDELGXIKLK	DXIXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
	201	XXXXAXXXXX	XXXXXXXXXS	GMKLTEALFQ	EPSFAYVNWS	AVKTADKDSQ
	251	WLKDVTEAYN	SDAFKAYAHK	RFEGYKSPAA	WNEGAAK*	

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence (SEQ ID NO: 221):

```
5      1  ATGAAAACCT TCTTCAAAAC CCTTTCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
     151  GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
     201  GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTGCGCC
     251  CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTCCAACAC
     301  AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCGA
     351  AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
     401  AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
     451  CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTGCGAT
     501  CAAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
     551  CCGAAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
     601  CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
     651  CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCTA GAACCGAGCT
     701  TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
     751  TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
     801  CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
     851  GCGCAGCCAA ATAA
```

This encodes a protein having amino acid sequence (SEQ ID NO: 222; ORF4a-1):

```
25      1  MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
      51  VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
     101  KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLSLEEVK DGSTVSAPND
     151  PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
     201  PRSRADVDFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
     251  WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*
```

ORF4a-1 (SEQ ID NO: 222) and ORF4-1 (SEQ ID NO: 218) show 99.7% identity in 287 aa overlap:

```
35      10      20      30      40      50      60
    orf4a-1  MKTFFKTL SAAALAL ILAACGG QKDSAPA ASASAAAD NGAAKKEI VFGTTVG DFGDMVKE
    orf4-1    MKTFFKTL SAAALAL ILAACGG QKDSAPA ASASAAAD NGAAKKEI VFGTTVG DFGDMVKE

      70      80      90     100     110     120
    orf4a-1  QIQPELE KGYTVKL VEFTDYV RPNLALA EGELDINV FQH KPYLDD FKKEHNL DITEVFQ
    orf4-1    QIQAELE KGYTVKL VEFTDYV RPNLALA EGELDINV FQH KPYLDD FKKEHNL DITEVFQ

     130     140     150     160     170     180
    orf4a-1  VPTAPLG LYPGKLS LEEVKDG STVSAPN DPSNFAR VLVMLDE LGWIKLK DGINPLT ASK
    orf4-1    VPTAPLG LYPGKLS LEEVKDG STVSAPN DPSNFAR VLVMLDE LGWIKLK DGINPLT ASK

     190     200     210     220     230     240
    orf4a-1  ADIAENL KNIKIVE LEAAQLP RSRADVD FAVVNGN YAISSGM KLTEALF QEPSFAY VNWS
```



```

      |||
orf4-1.  ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
           190       200       210       220       230       240

           250       260       270       280
5  orf4a-1  AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAX
      |||
orf4-1  AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAX
           250       260       270       280

```

Homology with an outer membrane protein of Pasteurella haemolítica (accession q08869) (SEQ ID

10 NO: 1126).

ORF4 (SEQ ID NO: 216) and this outer membrane protein (SEQ ID NO: 1126) show 33% aa identity in 91aa overlap:

```

           10       20
lip2.pasha  MNFKKLLGVALVSALALTACKDEKAQAP----
15          || | :| | || | :| | :| :|
ORF4        VXTPNPDGRTPCPSFLFETATTSGENMKTFFKTLSAAAL--ALILAACGFKKTARPPHPL
           110       120       130       140       150

           30       40       50       60       70       80
lip2.pasha. -ATTAKTENKAPLKVGVMTGPEAQMTEVAVKIAKEKYGLDVELVQFTEYTPNAALHSKD
20          : : : | : | : : | : : : : : || | | :| :| :| :| :| :| :| :
ORF4        LPPPTTARRKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGE
           160       170       180       190       200       210

           90       100      110      120      130      140
lip2.pasha  LDANAFQTVPYLEQEVKDRGYKLAIIGNTLVWPPIAAYSKKIKNISLKDGTVAIPNNAS
25          |
ORF4        L.....

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 (SEQ ID NO: 216) shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) (SEQ ID NO: 224) from *N. gonorrhoeae*:

```

           10       20       30
orf4nm.pep  MKTFFKTL SAAALALILAACGXQKDSAPAA
30          ||| ||| :| :| ||| ||| ||| |||
orf4ng      RANAVXTPNPDGRTPCLSFLFETATTSGENMKTFFKTLSTASLALILAACGGQKDSAPAA
           200       210       220       230       240       250

           40       50       60       70       80       89
orf4nm.pep  SASA-AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
35          || :| :| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf4ng      SAAAPSADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
           260       270       280       290       300       310

```

-212-

5 orf4nm.pep 90  
 EDEL  
 orf4ng EGELDINVFQHKPYLDDFKKEHNLDITEAFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPN  
 320 330 340 350 360 370

The complete length ORF4ng nucleotide sequence (SEQ ID NO: 223) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 224):

10 1 MKTFFKTLST ASLALILAAC GGQKDSAPAA SAAAPSDNG AAKKEIVFGT  
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ  
 101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN  
 151 DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ  
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS  
 15 251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYP A WNEGAAK\*

Further analysis revealed the complete length ORF4ng DNA sequence (SEQ ID NO: 225) to be:

20 1 atgAAAACCT TCTTCAAAAC cctttccgcc gccgcaCTCG CGCTCATCCT  
 51 CGCAGCCTGc ggCggtcaAA AAGACAGCGC GCCCgcagcc tctgcCGCCG  
 101 CCCCTTCTGC CGATAACGgc gCgGCGAAAA AAGAAAtcgt ctTCGGCAGC  
 151 Accgtgggcg acttcggcgA TAtggTCAAA GAACAAATCC AagcCGAgct  
 201 ggAGAAAAAA GgctACACcg tcAAattggt cgaatttacc gactatgtGC  
 251 gCCCCGAATCT GGCATTGGCG GAGGGCGAGT TGGACATCAA CGTCTTCCAA  
 301 CACAAACCTT ATCTTGACGA TTTCAAAAAA GAACACAACC TGGACATCAC  
 351 CGAAGCCTTC CAAGTGCCGA CCGCGCCTTT GGGACTGTAT CCGGGCAAAC  
 401 TGAATTCGCT GGAAGAAGTC AAAGACGGCA GCACCGTATC CGCGCCCAac  
 451 gACcctTCCA ACTTCGCACG CGCCTTGGTG ATGCTGAACG AACTGGGTTG  
 501 GATCAAACTC AAAGACGGCA TCAATCCGCT GACCGCATCC AAAGCCGACA  
 551 TCGCGGAAAA CCTGAAAAAC ATCAAAATCG TCGAGCTTGA AGCCGCACAA  
 601 CTGCCGCGCA GCCGCGCCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA  
 651 CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCTGTTC CAAGAGCCGA  
 701 GCTTTGCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC  
 751 CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC  
 801 CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG  
 851 AAGGCGCAGC CAAATAA

This encodes a protein having amino acid sequence (SEQ ID NO: 226; ORF4ng-1):

40 1 MKTFFKTL~~SA~~ AALALILAAC GGQKDSAPAA SAAAPSDNG AAKKEIVFGT  
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ  
 101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN  
 151 DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ  
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS  
 251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYP A WNEGAAK\*

This shows 97.6% identity in 288 aa overlap with ORF4-1 (SEQ ID NO: 218):

45 orf4-1.pep 10 20 30 40 50 59  
 MKTFFKTL~~SA~~AALALILAACGGQKDSAPAASASA-AADNGAAKKEIVFGTTVGDFGDMVK  
 orf4ng-1 MKTFFKTL~~SA~~AALALILAACGGQKDSAPAASAAAPSDNGAAKKEIVFGTTVGDFGDMVK  
 10 20 30 40 50 60  
 50 60 70 80 90 100 110 119

	orf4-1.pep	EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVF				
	orf4ng-1	EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEAF	70	80	90	100
5	orf4-1.pep	QVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTAS	120	130	140	150
	orf4ng-1	QVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS	130	140	150	160
10	orf4-1.pep	KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW	180	190	200	210
	orf4ng-1	KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW	190	200	210	220
15	orf4-1.pep	SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAX	240	250	260	270
	orf4ng-1	SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAX	250	260	270	280

20 In addition, orf4ng-1 (SEQ ID NO: 226) shows significant homology with an outer membrane protein (SEQ ID NO: 1126) from the database:

	ID	LIP2_PASHA	STANDARD;	PRT;	276 AA.
	AC	Q08869;			
25	DT	01-NOV-1995	(REL. 32, CREATED)		
	DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)		
	DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)		
	DE	28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR.			
	SCORES	Initl: 279	Initn: 416	Opt: 494	
	Smith-Waterman score: 494; 36.0% identity in 275 aa overlap				
30	orf4ng-1.pep	MKTFFKTL SAAAL--ALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDM	10	20	30
	lip2_pasha	MNFKKLLGVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPEAQM	10	20	30
35	orf4ng-1.pep	VKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITE	60	70	80
	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTQPNALHSDKLDANAFQTVPYLEQEVKDRGYKLAI	60	70	80
40	orf4ng-1.pep	AFQVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT	120	130	140
	lip2_pasha	IGNTLVWP IAAYSKKIKNISELKDGATVAIPNNASNTARALLLQAHGLLKLKDPKN-VF	120	130	140
45	orf4ng-1.pep	ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA	180	190	200
	lip2_pasha	ATENDI IENPKNIKIVQADTSL LTRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKDSP	180	190	200

	180	190	200	210	220	230
	240	250	260	270	280	289
orf4ng-1.pep	YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAXX					
5 lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSFQTEEVFQEALKLFNGGVVKGW					
	240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (SEQ ID NO: 218) (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 (SEQ ID NO: 218) is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1 (SEQ ID NO: 218).

### Example 27

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 227):

```

1   CCTCGTCGTC CTCGGCATGC TCCAGTTTCA AGGGGCGATT TACTCCAAGG
25  51   CGGTGGAACG TATGCTCGGC ACGGTCATCG GGCTGGGCGC GGGTTTGGGC
    101  GTTTTATGGC TGAACCAGCA TTATTTCAC  GGCAACCTCC TCTTCTACCT
    151  CACCGTCGGC ACGGCAAGCG CACTGGCCGG CTGGGCGGCG GTCGGCAAAA
    201  ACGGCTACGT CCCTmTGCTG GCAGGGCTGA CGATGTGTAT GTCATCGGC
    251  GACAACGGCA GCGAATGGCT CGACAGCGGA CTCATGCGCG CCATGAACGT
    301  CCTCATCGGC GyGGCCATCG CCATCGCCGC CGCCAAACTG CTGCCGCTGA
    351  AATCCACACT GATGTGGCGT TTCATGCTTG CCGACAACCT GGCCGACTGC
    401  AGCAAAATGA TTGCCGAAAT CAGCAACGGC AGGCGCATGA CCCGCGAACG
    451  CCTCGAGGAG AACATGGCGA AAATGCGCCA AATCAACGCA CGCATGGTCA
    501  AAAGCCGCAG CCATCTCGCC GCCACATCGG GCGAAAGCTG CATCAGCCCC

```

5 551 GCCATGATGG AAGCCATGCA GCACGCCCAC CGTAAATCG TCAACACCAC  
 601 CGAGCTGCTC CTGACCACCG CCGCCAAGCT GCAATCTCCC AAACCTCAACG  
 651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....  
 701 ..... GC AGACACGCCC GCCGCATCCG  
 751 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC  
 801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA  
 851 ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA  
 901 TGCCACAGAA CGCCAACACC TGCGCCAAAG CCTGCTTGA

10 This corresponds to the amino acid sequence (SEQ ID NO: 228; ORF8):

15 1 .....PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR  
 51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQT  
 101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AHDPRTPRGE HGENAPNORT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP  
 201 XNRQHHRAAP DHRRQAASQ TQRQNPAAAX PPLHTAPN... .....Q  
 251 TRPPHPRHRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH  
 301 PPQMAGCPRT PTPAPKPA\*

Computer analysis of this amino acid sequence gave the following results:

## 20 Sequence motifs

ORF8 (SEQ ID NO: 228) is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

## Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF8 (SEQ ID NO: 228) shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) (SEQ ID NO: 230) from *N. gonorrhoeae*:

30 orf8ng 1 MDRDDLRRPRHAPVPRDLLQGGGYARYGHRAGRGFGRFMAEPALFPR 50  
 ||||| | ||| |||||:|||||  
 orf8.pep 1 .....PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR 44  
 30 orf8ng 51 QPPLLPHRRHGKRTGRLGGGRQKRLRPYVGGADDVHAHRRQRQRMARQRP 100  
 ||||| | ||||| ||||| | ||||:|||||  
 orf8.pep 45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94  
 orf8ng 101 DARDERPHRRRHRHCRRQTAAAEIHTDVAFHACRQPGRLQNDNCRNQQRQ 150  
 || ||||| || ||||| ||||| |||||  
 35 orf8.pep 95 HARHERPHRRGHRHRRRQTAAAEIHTDVAFHACRQPGRMQNDNCRNQQRQ 144  
 orf8ng 151 AYDARTFGAEYGQNA PNQRT HGQKPQPSRRHIGRKLHQPLHDGSHAARPP 200  
 |:| | |:|:||||| ||||| ||| |||||  
 orf8.pep 145 AHDPRTPRGEHGENAPNQRT HGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194

```

orf8ng      201 QNRQHHRAAPDHRRQAAISQTQRQRNPAARPPPLHTAPNRPATNRRPHQRQ 250
              |||||
orf8.pep    195 XNRQHHRAAPDHRRQAAISQTQRQRNPAAXPPLHTAPN.....Q 244

orf8ng      251 TRPPHPHRHRHQPRGTGSPRRTPPLPMAGFPLAQHQYASGNFRPRHPPATH 300
              |||||
orf8.pep    245 TRPPHPHRHRHQPRGTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294

orf8ng      301 PPQMAGCPRTPTPAPKPA* 319
              |||||
orf8.pep    295 PPQMAGCPRTPTPAPKPA* 313

```

The complete length ORF8ng nucleotide sequence (SEQ ID NO: 229) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 230):

```

1  MDRDDRLRRP RHAPVPRDL LQGGTYARY GHRAGRGFGR FMAEPALFPR
51 QPPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP
101 DARDERPHRR RHRHCRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
151 AYDARTFGAE YGQNAPNQRT HGQKPQPPRR HIGRKPHQPL HDGSHAARPP
201 QNRQHHRAAP DHRRQAAISQ TQRQRNPAAR PPLHTAPNRP ATNRRPHQRQ
251 TRPPHPHRHR HQPRGTGSPR TPPLPMAGFP LAQHGYASGN FRPRHPPATH
301 PPQMAGCPRT PTPAPKPA*

```

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 28

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 231):

```

1  ..GAAATCAGCC TGCAGTCCGA CNACAGGCCG GTTTCCTGN CGAAGCGGCG
51 GGATTCGAA CGTTTCTGC TGTGACGG CGGCAACAGC CGGCTCAAGT
101 GGGCGTGGGT GGAACCGG ACGTTCGAA CCGTCGGTAG CGCGCGTAC
151 CGCGATTGT CGCCTTTGGG CGCGAGTGG GCGGAAAAGG CGGATGAAA
201 TGTCCGCATC GTCGTTGCG CTGTGTGCGG AGAATCAAA AAGGCACAAG
251 TGCAGGAACA GTCGCCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
301 GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
351 CCGCTGGTTC AACGCCTTGG GCAGCCGCG CTTCAGCCGC AACGCCTGCG
401 TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
451 GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
501 AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
551 GTTATCCTTT CCCGACCGG..

```

This corresponds to the amino acid sequence (SEQ ID NO: 232; ORF61):

```

1  ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAUVENG TFATVGSAPY
51 RDLSPGAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
101 AXGIRNHYRH PEEHGS DRWF NALGSRRFSR NACVVVSCGT AVTVDALTD
151 GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 233):

```

      1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
    51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
   101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
5      151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTTGGCGGT
   201  TTTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
   251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
   301  GCGCGGATTG CGCCGGACAA GCGGCACAAA ACCATATGCG TGACCCACCT
   351  GCAAAGTAAG GGCAGGGGCG GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
  10      401  GCGAGTGTCT GATGTTTCAGT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
   451  GAGTTGGGTT CGCTGTCGCC TGTTCGGCA GTGGCGTGTC GGC CGCCTT
   501  GTCGCGTTTA GGTTCGATG TGCAGATTAA GTGGCCCAAT GATTTGGTTG
   551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
   601  GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
  15      651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
   701  GGCAGGGCAA TGCCGATGCC GCCGTGCTGC TGGAACCGCT GTTGGTGAA
   751  CTGGACGCGG TGTGTTGCA ATATGCGCGG GACGGATTG CGCCTTTTGT
   801  GGC GGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
   851  TGC GCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
  20      901  CAAGGCGTTT TGCACCTGGA AACGGCAGAG GGCAAAACAGA CGGTCGTCAG
   951  CCGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
  1001  GGC GGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
  1051  AAGTGGGCGT GGGTGAAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
  1101  GTACCGCGAT TTGTCGCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
  25      1151  GAAATGTCCG CATCGTCGGT TGC GCTGTGT GCGGAGAATT CAAAAGGCA
   1201  CAAGTG CAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
   1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
   1301  CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
   1351  TGCGTGCTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
  30      1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
   1451  AAGAATCGCT CGCCGTCGCA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
   1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
   1551  GGATGCGGTT TCGCGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
  35      1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
   1651  GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTGTGCGG AAAATACCGT
   1701  GCGCGTGGCG GACAACCTCG TCATTACGG GTTGTGAAC ATGATTGCCG
  1751  CCGAAGGCAG GGAATATGAA CATATTTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 234; ORF61-1):

```

40      1  MTVLKLSHWR VLAELADGLP QHVSQ LARMA DMKPQQLNGF WQQMPAHIRG
    51  LLRQHDGYWR LVRPLAVFDA EGLREL GERS GFQTALKHEC ASSNDEILEL
   101  ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFD RPQY
   151  ELGSLSPVAA VACRRALSRL GLDVQIKWPN DLVVGRDKLG GILIETVRTG
   201  GKTVA VVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLVE
45      251  LDAVLLQYAR DGFAPFVAEY QANRDHGKA VLLLRDGETV FEGTVKGV DG
   301  QGV LHLETA E GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
   351  KWAWENGTF ATVGSA PYRD LSPLGA EWA E KADGNVRIVG CAVCGEFKKA
   401  QVQEQLARKI EWL PSSA QAL GIRNHYRHPE EHGS DRWFNA LGSRRFSRNA
   451  CVV VSCGTAV TVDALTDG H YLGGT IMPGF HLMKESLAVR TANLNRHAGK
50      501  RYPFPTTTGN AVASGMMDAV CGSVM MMHGR LKEKTGAGKP VDVII TGGGA
   551  AKVAEALPPA FLAENTVRVA DNLVIYGLLN MIAAEGREYE HI*

```

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1 (SEQ ID NO: 234). Further computer analysis of this amino acid sequence gave the following results:

Homology with the baf protein of *B. pertussis* (accession number U12020) (SEQ ID NO: 1127).

ORF61 (SEQ ID NO: 232) and baf protein (SEQ ID NO: 1127) show 33% aa identity in 166aa overlap:

```

5      orf61  23  LLLDGGNSRLKAWWVE-NGTFATVGSAPYR----DLSPDLGAWEAKADGNVRIVGCAVCG 77
      baf    3   +L+D GNSRLK W + + A AP DL LG A R +G V G
      orf61  78  EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNALGSRRFSRN 131
      baf    63  + + L I WL + A G+RN YR+P++ G+DRW L +
      orf61  132 ACVVVSCGTAVTVDALTTDDGHYLGXGTIMPGFHLMKESLAVRTANL 177
      baf    123 +V S GTA T+D + D + G G I+PG +M+ +LA TA+L
      baf    123 PLLVASFGTATTLDITIGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF61 (SEQ ID NO: 232) shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) (SEQ ID NO: 236) from strain A of *N. meningitidis*:

```

20      orf61.pep                                10      20      30
      orf61a      TVFEGTVKGVDDGQGVHLHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSERFLLLDGGNS
      290      300      310      320      330      340

      orf61.pep      40      50      60      70      80      90
      orf61a      RLKAWVNGTTFATVGSAPYRDLSPDLGAWEAKADGNVRIVGCAVCGEFKKAQVQEQLAR
      350      360      370      380      390      400

      orf61.pep      100      110      120      130      140      150
      orf61a      KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTTDD
      410      420      430      440      450      460

      orf61.pep      160      170      180      189
      orf61a      GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT
      470      480      490      500      510      520

      orf61a      HGRLKEKTGAGKPDVVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG
      530      540      550      560      570      580

```

The complete length ORF61a nucleotide sequence (SEQ ID NO: 235) is:

```

40      1  ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA

```



```

51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
201 TTTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
5   251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGACAA GGCACACAAA ACCATATGTG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGTCT GATGTTTCACT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGCC GGCAGCCTT
10  501 GTCGCGTTTG GGTTTGAAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
601 GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTC TGCTGCCCAA
651 GGAAGTGGA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GGCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
15  751 CTTGATGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTC CGCCTTTTGT
801 GGCAGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
901 CAAGCGGTTT TGCATTGGA AACGGCAGAG GGCAAAACAGA CGGTCGTCAG
951 CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGTTTTC GTGCCGAAGC
20  1001 GGCGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
1101 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGTGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATT CAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
25  1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGACGCC GCCGTTTCAG CCGCAACGCC
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
30  1501 CGTTATCCTT TCCCGACAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
1551 GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCAAAAGTTG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACCGG GCTGCTGAAC CTGATTGCCG
35  1751 CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 236):

```

1  MTVLKPSHWR VLAELADGLP QHVSQARMAD MKPQQLNGF WQQMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLRELTERS GFQTALKHEC ASSNDEILEL
40  101  ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDPRQY
151  ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGDRDLG GILLETVRTG
201  GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251  LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVVG
45  301  QGVLEHLEAE GKQTVVSGEI SLRSDRPVS VPKRRDSERF LLLDGGNSRL
351  KWAVVENGTG ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
401  QVQEQALARKI EWLPSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFNRNA
451  CVVVSCGTAV TVDALDDGH YLGGTIMPGF HLMKESLAVR TANLNRHAGK
501  RYFPPTTGN AVASGMMDAV CGSVMMHGR LKEKTGAGKP VDVITGGGA
50  551  AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a (SEQ ID NO: 236) and ORF61-1 (SEQ ID NO: 234) show 98.5% identity in 591 aa overlap:

```

55  orf61a.pep      10      20      30      40      50      60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf61-1      MTVLKPSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR

```

		10	20	30	40	50	60
		70	80	90	100	110	120
	orf61a.pep	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK					
5	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf61a.pep	GRGRQGRKWSHRLGECLMFSGWVFDPRQYELGSLSPVAACRRALSRLGLKTQIKWPN					
10	orf61-1	GRGRQGRKWSHRLGECLMFSGWVFDPRQYELGSLSPVAACRRALSRLGLDVQIKWPN					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf61a.pep	DLVVGRDKLGGILITVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA					
15	orf61-1	DLVVGRDKLGGILITVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVGDG					
20	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVGDG					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf61a.pep	QGV LHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGNSRLKWAVVNGTF					
25	orf61-1	QGV LHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGNSRLKWAVVNGTF					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf61a.pep	ATVGSAPYRDLSP LGAEWA EKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL					
30	orf61-1	ATVGSAPYRDLSP LGAEWA EKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL					
		370	380	390	400	410	420
		430	440	450	460	470	480
	orf61a.pep	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHYLGGTIMPGF					
35	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHYLGGTIMPGF					
		430	440	450	460	470	480
		490	500	510	520	530	540
	orf61a.pep	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMHGR LKEKTGAGKP					
40	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMHGR LKEKTGAGKP					
		490	500	510	520	530	540
		550	560	570	580	590	
	orf61a.pep	VDV IITGGGA AKVAEALPPAFLAENTVRVADNLVIHGLLN LIAAEGGESEHTX					
45	orf61-1	VDV IITGGGA AKVAEALPPAFLAENTVRVADNLVIYGLLN MIAAEGREYEHIX					
		550	560	570	580	590	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 (SEQ ID NO: 232) shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) (SEQ ID NO: 238) from *N. gonorrhoeae*:

	orf61.pep	EISLRSDXRPVSVXKRRDSEFLLLDGGNS	30
5	orf61ng	TVCEGTVKGV DGRGV LHLETAEGEQTVVSGEISLRPDNRSVSVKRPDSEFLLLEGGNS	211
	orf61.pep	RLKWAVVENGT FATVGSAPYRDL SPLGAEWA EKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAVVENGT FATVGSAPYRDL SPLGAEWA EKADGNVRIVGCAVCGESKKAQVKEQLAR	271
10	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTTDD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTTDD	331
	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYFPPT	189
15	orf61ng	GHYLG-GTIMPGFHLMKESLAVRTANLNRPA GKRYFPPTTTGNAVASGMMDAVCGSIMMM	390

An ORF61ng nucleotide sequence (SEQ ID NO: 237) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 238):

	1	MFSFGWAFDR	PQYELGSLSP	VAALACRRAL	GCLGLETQIK	WPNDLVVGRD
	51	KLGGILIE TV	RAGGKTVA VV	GIGINFVLPK	EVENAASVQS	LFQTASRRGN
20	101	ADAAVLLE TL	LAELGAVLE Q	YAEEGFAPFL	NEYETANRDH	GKAVLLLRDG
	151	ETVCEGTVKG	VDGRGV LHLE	TAEGEQTVVS	GEISLRPDNR	SVSVKRPDS
	201	ERFLLLEGGN	SRLKWAVVEN	GT FATVGSAP	YRDL SPLGAE	WAEKADGNVR
	251	IVGCAVCGES	KKAQVKEQLA	RKIEWLPSSA	QALGIRNHYR	HPEEHGSDRW
	301	FNALGSRRFS	RNACVVVSCG	TAVTVDALTD	DGHYLG GTIM	PGFHLMKESL
25	351	AVRTANLNRP	AGKRYFPPTT	TGNAVASGMM	DAVCGSIMMM	HGRLKEKNGA
	401	GKPV DVIITG	GGAAKVAEAL	PPAFLAENTV	RVADNLVIHG	LLNLIAAEGG
	451	ESEHA*				

Further analysis revealed the complete gonococcal DNA sequence (SEQ ID NO: 239) to be:

30	1	ATGACGGTTT	TGAAGCCTTC	GCATTGGCGG	GTGTTGGCGG	AGCTTGCCGA
	51	CGGTTTGCCG	CAACACGTAT	CGCAATTGGC	GCGTGAGGCG	GACATGAAGC
	101	CGCAGCAGCT	CAACGGTTTT	TGGCAGCAGA	TGCCGCGCGA	TATACGCGGG
	151	CTGTTGCGCC	AACACGACGG	CTATTGGCGG	CTGGTGCGCC	CCTTGCGCGT
	201	TTTCGATGCC	GAAGGTTTGC	GCGATCTGGG	GGAAAGGTCG	GGTTTTCAGA
35	251	CGGCATTGAA	GCACGAGTGC	GCGTCCAGCA	ACGACGAGAT	ACTGGAATTG
	301	GCGCGGATTG	CGCCGACAAA	GGCGCACAAA	ACCATATGCG	TGACCCACCT
	351	GCAAAGTAAG	GGCAGGGGGC	GGCAGGGGCG	GAAGTGGTCG	CACCGTTTGG
	401	GCGAGTGCCT	GATGTT CAGT	TTCGGCTGGG	CGTTTGACCG	GCCG CAGTAT
	451	GAGTTGGGTT	CGCTGTGCGC	TGTTGCGGCA	CTTGCGTGCC	GGCGCGCTTT
40	501	GGGGTGTTTG	GGTTTGGA AA	CGCAAATCAA	GTGGCCAAAC	GATTTGGTCG
	551	TCGGACGCGA	CAAATTTGGC	GGCATTCTGA	TTGAAACAGT	CAGGGCGGGC
	601	GGTAAACGCG	TTGCCGTGGT	CGGTATCGGC	ATCAATTTCG	TGCTGCCCCA
	651	GGAAGTGGA A	AACGCCGCTT	CCGTGCAGTC	GCTGTTTCAG	ACGGCATCGC
	701	GGCGGGGCAA	TGCCGATGCC	GCCGTATTGC	TGGAAACATT	GCTTGCGGAA
45	751	CTGGGCGCGG	TGTTGGAACA	ATATGCGGAA	GAAGGGTTCG	CGCCATTTT
	801	AAATGAGTAT	GAAACGGCCA	ACCGCGACCA	CGGCAAGGCG	GTATTGCTGT
	851	TGCGCGACGG	CGAAACCGTG	TGCGAAGGCA	CGGTTAAAGG	CGTGGACGGA
	901	CGAGGCGTTC	TGCACTTGA A	AACGGCAGaa	ggcgaACAGa	cggtcgtcag
	951	cggcgaaaTC	AGcctGCggc	ccgacaacaG	GTCGGTttcc	gtgccgaagc

5	1001	ggcccgatTC	GgaacgtTTT	tTGctgttg	aaggcgggaa	cagccgGCTC
	1051	AAGTGGGCGT	GggtggAAAa	cggcacgttc	gcaaccgtgg	gcagcgcgCc
	1101	gtaCCGCGAT	TTGTGCCTT	TGGGCGCGGA	GTGGGCGGAA	AAGGCGGATG
	1151	GAAATGTCCG	CATCGTCGGT	TGCGCCGTGT	GCGGAGAATC	CAAAAAGGCA
	1201	CAAGTGAAGG	AACAGCTCGC	CCGAAAAATC	GAGTGGCTGC	CGTCTTCCGC
10	1251	ACAGGCTTTG	GGCATAACGA	ACCACTACCG	CCACCCCGAA	GAACACGGTT
	1301	CCGACCGTTG	GTTCAACGCC	TTGGGACGCC	CCGCTTTCAG	CCGCAACGCC
	1351	TGCGTCTGTC	TCAGTTGCGG	CACGGCGGTA	ACGGTTGACG	CGCTCACCGA
	1401	TGACGGACAT	TATCTCGGCG	GAACCATCAT	GCCCGGCTTC	CACCTGATGA
	1451	AAGAATCGCT	CGCCGTCCGA	ACCGCAACC	TCAACCGCCC	CGCCGGCAAA
15	1501	CGTTACCTTT	TCCCGACCAC	AACGGGCAAC	GCCGTGCGAA	GCGGCATGAT
	1551	GGACGCGGTT	TGCGGCTCGA	TAATGATGAT	GCACGGCCGT	TTGAAAGAAA
	1601	AAAACGGCGC	GGGCAAGCCT	GTCTGATGTA	TCATTACCGG	CGGCGGCGCG
	1651	GCGAAAGTCG	CCGAAGCCCT	GCCGCTGCA	TTTTTGGCGG	AAAAATCCGT
	1701	GCGCTGGGCG	GACAACCTCG	TCATCCACGG	GCTGCTGAAC	CTGATTGCCG
	1751	CCGAAGCGCG	GGAATCGGAA	CACGCTTAA		

This corresponds to the amino acid sequence (SEQ ID NO: 240; ORF61ng-1):

20	1	MTVLKPSHWR	VLAELADGLP	QHVSQALAREA	DMKPQQLNGF	WQQMPAHIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRDLGERS	GFQTALKHEC	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HRLGECLMFS	FGWAFDRPQY
	151	ELGSLSPVAA	LACRRALGCL	GLETQIKWPN	DLVVGRDKLG	GILIETVRAG
25	201	GKTVAVVVIG	INFVLPEKEV	NAASVQSLFQ	TASRRGNADA	AVLLETVLAE
	251	LGAVLEQYAE	EGFAPFLNEY	ETANRDHGKA	VLLLRDGETV	CEGTVKGVVG
	301	RGVLHLETAE	GEQTVVSGEI	SLRPDNRSVS	VPKRPDSERF	LLLEGGNSRL
	351	KWAVVENGTF	ATVGSAPYRD	LSPLGAEWAE	KADGNVRIVG	CAVCGESKKA
30	401	QVKEQLARKI	EWLPSSAQAL	GIRNHRYRHE	EHGSDRWFNA	LGSRFRFSRNA
	451	CVVVSCGTAV	TVDALTTDDGH	YLGGTIMPGF	HLMKESLAVR	TANLNRPAKG
	501	RYFPPTTTGN	AVASGMMDAV	CGSIMMMHGR	LKEKGAGAKP	VDVITITGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIHGLLN	LIAAEGGESE	HA*

ORF61ng-1 (SEQ ID NO: 240) and ORF61-1 (SEQ ID NO: 234) show 93.9% identity in 591 aa overlap:

35	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQRLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	60
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQRLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	60
	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTCVTHLQSK	120
	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTCVTHLQSK	120
40	orf61ng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN	180
	orf61ng-1.pep	DLVVGRDKLGGILIIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILIIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
45	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEFGFAPFLNEYETANRDHGKAVLLLLRDGETVCEGTVKGVDG	300
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLLRDGETVFEGTVKGVDG	300
50	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF	360
		:	

```

orf61-1      QGVHLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGSRLKWAVVENGTF 360

orf61ng-1.pep ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420
|||||
orf61-1      ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQQLARKIEWLPSSAQAL 420

5  orf61ng-1.pep GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF 480
|||||
orf61-1      GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF 480

orf61ng-1.pep HLMKESLAVRTANLNRPAKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540
|||||
10 orf61-1      HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

orf61ng-1.pep VDVIIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLI AEGGESEHAX 593
|||||
orf61-1      VDVIIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMI AEGGREYEHIX 593

```

Based on this analysis, including the homology with the baf protein (SEQ ID NO: 1127) of *B. pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 29

The following partial DNA sequence was identified in *N. meningitidis* (SEQ ID NO: 241):

```

20      1  ATGTTTTACC  AAATCCTTGC  CCTGATTATC  TGGAGCAGCT  CGTTTATTGC
      51  CGCCAAATAT  GTCTATGGCG  GCATCGATCC  CGCATTGATG  GTCGGCGTGC
     101  GCCTGCTAAT  TGCCGCGCTG  CCTGCACTGC  CCGCCTGCCG  CCGTCATGTC
     151  GGCAAGATTC  CGCGTGAGGA  ATGGAAGCCG  TTGCTGATTG  TGTCGTTCTG
     201  CAACTATGTG  CTGACCCTGC  TGCTTCAGTT  TGTCGGGTG  AAATACACTT
     25  251  CCGCCGCCAG  CGCATCGGTC  ATTGTCGGAC  TCGAGCCGCT  GCTGATGGTG
     301  FTTGTCCGAC  ACTTTTTCTT  CAACGACAAA  GCGCGTGCC  ACCACTGGAT
     351  ATGCGGCGCG  GCGGCATTG  CCGGTGTCGC  GCTGCTGATG  GCGGGCGGTG
     401  CGGaAGAGGG  CGGCGaAGTC  GGCTGGTTCG  GCTGCCTGCT  GGTGTTGTG
     451  GCGGGCGCGG  GCTTTTGTGC  CGCTATGCGT  CCGACGCAAA  GGCTGATTGC
     30  501  ACGCATCGGC  GCACCGGCAT  TCACATCTGT  TTCCATTGCC  GCCGCATCGT
     551  TGATGTGCCT  GCCGTTTTCG  CTTGCTTTGG  CGCAAAGTTA  TACCGTGGAC
     601  TGGAGCGTCG  GGATGGTATT  GTCGCTGCTG  TATTTGGGTT  TGGGGTGC..

```

This corresponds to the amino acid sequence (SEQ ID NO: 242; ORF62):

```

35      1  MFYQILALII  WSSSFIAAKY  VYGGIDPALM  VGVRLLIAAL  PALPACRRHV
      51  GKIPREEWKP  LLIVSFVNYV  LTLLQLFVGL  KYTSAASASV  IVGLEPLLMV
     101  FVGHHFFNDK  ARAYHWICGA  AAFAGVALLM  AGGAEEGGEV  GWFGCLLVLL
     151  AGAGFCAAMR  PTQRLIARIG  APAFTSVSIA  AASLMCLPFS  LALAQSYTVD
     201  WSVGMLVSL  YLGLGC..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 243):

```

1  ATGTTTTACC  AAATCCTTGC  CCTGATTATC  TGGAGCAGCT  CGTTTATTGC

```

51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC  
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC  
 151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCTG  
 201 CAACTATGTG CTGACCTGTC TGCTTCAGTT TGTGCGGTTG AAATACACTT  
 5 251 CCGCCGCCAG CGCATCGGTC ATTGTGCGAC TCGAGCCGCT GCTGATGGTG  
 301 TTTGTGCGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT  
 351 ATGCGGCGCG GCGGCATTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG  
 401 CGGAAGAGGG CCGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTGTGTTG  
 451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC  
 10 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT  
 551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC  
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTGGGTT TGGGGTGC GG  
 651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCTGCCA  
 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG  
 15 751 GCGGTTTGA TTTTGGGCGA ACACCTGTG CCGTGTCGG CCTTGGGCGT  
 801 GTTTGTCGTC ATCGCCGCCA CCTTGTTGTC GCGCCGCTG TCGATCAAA  
 851 AATAA

This corresponds to the amino acid sequence (SEQ ID NO: 244; ORF62-1):

20 1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV  
 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV  
 101 FVGHFFFNK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL  
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD  
 201 WSVGMVLSLL YLGLGCGWYA YWLWNKMSR VPANVSGLLI SLEPVVGVL  
 251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147) (SEQ ID NO: 1128)

30 ORF62 (SEQ ID NO: 242) and HI0976 (SEQ ID NO: 1128) show 50% aa identity in 114aa overlap:

Orf62 1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXCRRHVKGKIPREEWKP 60  
 M YQILAL+IWSSS I K Y +DP L+V VR R KI + K  
 HI0976 1 MLYQILALLIWSSSLIVGKLTYSMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60  
 35 Orf62 61 LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNKARAY 114  
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +  
 HI0976 61 LWLAFPNYTAFLQLQFGLKYTSASSAVTMIGLEPLLVVVGHFFFKTKQNGF 114

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF62 (SEQ ID NO: 242) shows 99.5% identity over a 216aa overlap with an ORF (ORF62a)  
 40 (SEQ ID NO: 246) from strain A of *N. meningitidis*:

-225-

		10	20	30	40	50	60
	orf62.pep	<u>MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV</u> GKIPREEWKP					
	orf62a	<u>MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHV</u> GKIPREEWKP					
5		10	20	30	40	50	60
	orf62.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
	orf62a	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
10		70	80	90	100	110	120
	orf62.pep	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
	orf62a	LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
		70	80	90	100	110	120
	orf62.pep	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA					
	orf62a	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA					
15		130	140	150	160	170	180
	orf62.pep	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA					
	orf62a	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA					
		130	140	150	160	170	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC					
	orf62a	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI					
20		190	200	210	220	230	240
	orf62a	SLEPVVGVLAVLILGEHLSPVSVLGVFVVIATLVAGRLSHQKX					
		250	260	270	280		

The complete length ORF62a nucleotide sequence (SEQ ID NO: 245) is:

25	1	ATGTTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGECAAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
	101	GCCTGCTGAT	TGCTGCGCTG	CCTGCACTGC	CCGCGTGCCG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCTG
	201	CAACTATGTG	CTGACCCTGC	TACTTCAGTT	TGTCGGGTTG	AAATACACTT
30	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCACT	GCTGATGGTG
	301	TTTGTCTGGAC	ACTTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTTC	CCGGTGTCTG	GCTGCTGATG	GCGGGCGGGT
	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTTC	GCTGCCTGCT	GGTGTGTTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
35	501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTTC	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GAATGGTATT	GTCGCTGCTG	TATTTGGGCG	TGGGGTGCAG
	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCTTGCCA
	701	ACGTTTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGCTGCTG
40	751	GCGGTTTTGA	TTTTGGGCCA	ACACCTGTCG	CCCGTGTCGG	TCTTGGGCGT
	801	GTTTGTCTGC	ATCGCCGCCA	CCTTGGTTGC	CGGCCGGCTG	TCGCATCAAA
	851	AATAA				

This encodes a protein having amino acid sequence (SEQ ID NO: 246):

45	1	MFYQILALII	WSSSFIAAKY	VYGGIDPALM	VGVRLLIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLLLQFVGL	KYTSASASV	IVGLEPLLMV
	101	FVGHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
	151	AGAGFCAAMR	PTQRLIARIG	APAFTSVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSLL	YLVGCSWYA	YWLWNKGMSR	VPANVSGLLI	SLEPVVGVL
50	251	AVLILGEHLS	PVSVLGVFVV	IAATLVAGRL	SHQK*	

ORF62a (SEQ ID NO: 246) and ORF62-1 (SEQ ID NO: 244) show 98.9% identity in 284 aa overlap:

5	orf62a.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP	60
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP	60
10	orf62a.pep	LLIVSFVNYVLTLLQFVGLKYSAAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
	orf62-1	LLIVSFVNYVLTLLQFVGLKYSAAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
15	orf62a.pep	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62-1	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62a.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI	240
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI	240
	orf62a.pep	SLEPVVGVLAVLILGEHLSPVSVLGVFVVAATLVAGRLSHQXX	285
	orf62-1	SLEPVVGVLAVLILGEHLSPVSALGVFVVAATLVAGRLSHQXX	285

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 (SEQ ID NO: 242) shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) (SEQ ID NO: 248) from *N. gonorrhoeae*:

25	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP	60
	orf62.pep	LLIVSFVNYVLTLLQFVGLKYSAAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLQFVGLKYSAAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
	orf62.pep	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
30	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	216
	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI	240

The complete length ORF62ng nucleotide sequence (SEQ ID NO: 247) is:

35	1	ATGTTTACC	AAATCCTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCCGCGTGC
40	101	GCCTGCTGAT	TGCCGCGCTG	CCTGCACTGC	CCGCCTGCCG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCGT
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTGCGAC	ACTTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT



5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85

351 ATGCGGCGCG GCGGCATTTC CCGGTGTCGC GCTGCTGATG GCGGGCGGTG  
401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG  
451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAA GGCTGATTGC  
501 CCGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT  
551 TGATGTGCCT GCCGTTTTTC CTTGCTTTGG CGCAAAGTTA TACCGTGGAC  
601 TGGAGCGTCG GGATGGTATT GTCGCTGTTG TATTGGGTG TGGGGTGCGG  
651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA  
701 ACGCGTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGTTG  
751 GCGGTTTTGA TTTGGGCGA ACATTATCG CCCGTGTCCG CTTGGGCGT  
801 GTTTGTCGTC ATCGCCGCCA CTTTCGCCGC CGGCCGGCTG TCGCGCAGGG  
851 ACGCGCAAAA CGGCAATGCC GTCTGA

This encodes a protein having amino acid sequence (SEQ ID NO: 248):

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

1 MFYQILALII WGSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV  
51 GKIPREEWKP LLIVSFVNYV LTLQLQFVGL KYTSAASASV IVGLEPLLMV  
101 FVGHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL  
151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD  
201 WSVGMLVSL LYLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVL  
251 AVLILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V\*

ORF62ng (SEQ ID NO: 248) and ORF62-1 (SEQ ID NO: 244) show 97.9% identity in 283 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995  
1000

orf62ng.pep MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV  
orf62-1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIIAALPALPACRRHV  
orf62ng.pep LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMV  
orf62-1 LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMV  
orf62ng.pep AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAF  
orf62-1 AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAF  
orf62ng.pep AASLMCLPFSLALAQSYTVDWSVGMLVSLLYLGLGCGWYAYWLWNKGMSRV  
orf62-1 AASLMCLPFSLALAQSYTVDWSVGMLVSLLYLGLGCGWYAYWLWNKGMSRV  
orf62ng.pep SLEPVVGVLAVLILGEHLSPVSAALGVFVVIAATFAAGRLSRRDAQNGNA  
orf62-1 SLEPVVGVLAVLILGEHLSPVSAALGVFVVIAATLVAGRLSHQKX

Furthermore, ORF62ng (SEQ ID NO: 248) shows significant homology to a hypothetical  
50 *H. influenzae* protein (SEQ ID NO: 1128):

sp|Q57147|Y976\_HAEIN HYPOTHETICAL PROTEIN HI0976 )gi|1074589|pir||B64163  
 hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)  
 )gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128  
 Score = 106 bits (262), Expect = 2e-22  
 Identities = 56/114 (49%), Positives = 68/114 (59%)

Query: 1 MFYQILALI IWSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXXXCRRHVGVKIPREEWKP 60  
 M YQILAL+IW SS I K Y +DP L+V VR R KI + K  
 Sbjct: 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIAMIIVMPLFLRRWKKIDKPMRKQ 60

Query: 61 LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLL MVFVGHFFFN DKARAY 114  
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +  
 Sbjct: 61 LWLWLAFFNYTAVFLLQLQFGLKYTSAASAVTMIGLEPLL VVFGHFFFKTKQNGF 114

Based on this analysis, including the homology with the transmembrane protein (SEQ ID NO: 1128) of *H.influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 30

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 249):

```

1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA
51  sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
101 GGTGGATTGT TGC GTTCAGC GCAATGCTGC TGCTGGTGTG GTCCGCGGTT
151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
201 CGGTTTCGCTa srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
251 GCCGkACTGC CCGGCGTGTG TCTGTTCGGC TTTCCCGCAC AGTTCATCAA
301 CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
25 351 GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
401 CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
451 GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGCG AGCGGTTTTG
501 CCCGCTTGC CCTGTACAy ksCGCAAGCG GCAAAATCGA AAAAAGCATC
551 AACC CGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
30 601 AATCCaACGG GCGGGTTCGG TCAGGGATTT GGAAAGCATA GCGGCGGTAT
651 TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
701 GCCTTGTTTT TCCGTCAGCC GGTTCCTCAA GGCGTGGCAG AGGATGCCGT
751 yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
801 AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
35 851 TCGATTTTTTc TTGCACTGGT CATGGCACTG TATTTGCCCC GCCGTTTTCGT
901 CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
951 ATTTTCAGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
1001 ACCArGTTGT TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
1051 AGACGAGCGC AACCGCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
40 1101 GCGTGTTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAGGC
1151 TGTCTGAAAA CCTTCAACAA AGCGGCGGGT ACC..

```

This corresponds to the amino acid sequence (SEQ ID NO: 250; ORF64):

```

1  MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51  LARYVILLK DRRDGVFGSX XAKXPXXMF TLVAXLPGVF LFGFPAQFIN

```

5

```

101 GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVVPV QIDLIGAASL
151 PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PPGKARWEK
201 IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
251 LIEKARAKYA ELSYSKGLQ TFFLATLLIA SLLSIFLALV MALYFARRFV
301 EPVLSLAEGA KAVAQGDFSQ TRPVLRNDEF GRLTXLFNHM TEQLSIAKDA
351 DERNRRREEA ARHYLECVLE GLTTGVVVF D EQGCLKTFNK AAGT..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 251):

10

15

20

25

30

35

40

45

50

```

1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
51  CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
101 GGTGGATTGT TGCCTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
201 CGGTTTCGAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
251 TACTGCCCGG CGTGTTCCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
351 CCTCAATTG AGCAAGTCCG CATTGAATTT GGCGGCAGAC AACGCCCTCG
401 GCAACGCCGT CCCCGTGCAG ATAGACCTCA TCGGCGCGGC TTCCCTGCCC
451 GGGGATATGG GCAGGGTGCT GGAACATTAC GCCGCGAGCG GTTTTGCCCA
501 GCTTGCCCTG TACATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
551 CGCACAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC
601 CAACGGGCGG GTTCGGTCAG GGATTGGAA AGCATAGGCG GCGTATTGTA
651 CGCGCAGGGC TGGCTGTCGG CGGGTACGCA CAACGGGCGC GATTACGCCT
701 TGT'TTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
751 ATCGAAAAGG CAAGGCGGAA ATATGCTGAG TTGAGTTACA GCAAAAAGG
801 TTTGCAGACC TTTTCTCTGG CAACCTGCT GATTGCCTCG CTGCTGTCGA
851 TTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCCGCCG TTTCTGTCGA
901 CCCGTCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGCGGATTT
951 CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGCGCGCA GGAAGCCGCC AGGCATTATC TTGAATGCGT
1101 GTTGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGCTGTC
1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGCGC GCAGGTACGG
1301 ACAAAACGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCG CAATTATGCG
1651 CGTTCCCTT CGCTCAAATT GGAAAATCAG GATTGAAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
1751 TTGCCGCGCA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCCGATTG
1901 TCCTGACGGT TTGCGACAAC GGCAAAGGGT TCGGCAGGGA AATGCTGCAC
1951 AACCGCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
2001 TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
2051 TGAGCAATCA GGATGCGGGT GGCGCGTGTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAAAAA CTTATGCGTA G

```

This corresponds to the amino acid sequence (SEQ ID NO: 252; ORF64-1):

55

```

1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51  LARYVILLLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVVPQ IDLIGAASLP
151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI

```

-230-

5 201 QRAGSVRDLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPV PKGVAEDAVL  
 251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE  
 301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD  
 351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT  
 401 PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL  
 451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT  
 501 PIQLSAERLA WKLGGKLDEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA  
 551 RSPSLKLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRQ  
 10 601 VLHNIFKNAA EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFGREMLH  
 651 NAFEPYVTDK PAGTGLGLPV VKKIIEEHGG RISLSNQDAG GACVRIILPK  
 701 TVKTYA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF64 (SEQ ID NO: 250) shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) (SEQ ID NO: 254) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf64.pep	MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK					
20	orf64a	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK					
		10	20	30	40	50	60
	orf64.pep	DRRDGVFGSXXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN					
25	orf64a	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN					
		70	80	90	100	110	120
	orf64.pep	LSKSALNLAADNALGNAVVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE					
30	orf64a	LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE					
		120	130	140	150	160	170
	orf64.pep	KSINPHKLDQPPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP					
35	orf64a	KSINPHKLDQPPGKARWEKIQQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQP					
		180	190	200	210	220	230
	orf64.pep	VPKGVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV					
40	orf64a	VPKGVAEDAVLIEKARAXXXLSYSKKG LQTFFLATLLIASLLSIFLALVMALYFARRFV					
		240	250	260	270	280	290
	orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTXLFNHMT EQLSIAKDADERNRRREEA					
45	orf64a	EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEA					
		300	310	320	330	340	350

-231-

```

              370      380      390
orf64.pep    ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT
              |||||
orf64a       ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAEEQILGMPLTPLWGSSRHGWHGVSAQQSL
5            360      370      380      390      400      410

orf64a       LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQ
            420      430      440      450      460      470

```

The complete length ORF64a nucleotide sequence (SEQ ID NO: 253) is:

```

10      1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TCGCGCCGTCG TCCTGTTGTA
      51  CGGACTGACG GCGGCAACCG GCAGACCAG TTCGCTGGCG GATTATTTCT
     101  GGTGGATTGT TCGGTTACAG GCAATGCTGC TGCTGGTGTT GTCCGCGGTT
     151  TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
     201  CGGTTTCGAG ATTGCCAAAC GCCTTTCCGG GATGTTTACG CTGGTTGCCG
15      251  TACTGCCCGG CGTGGTTCTG TTCGGCGTTT CCGCACAGTT TATCAACGGC
     301  ACGATTAATT CGTGGTTTCG CAACGATACC CACGAGGCGC TTGAACGCAG
     351  CCTCAATTG AGCAAGTCCG CATTGAATCT GCGGCGAGAC AACGCCCTTG
     401  GCAACGCCAT CCGCGTGCAG ATAGACNTCA TCGGCGCGGC TTCCCTGCCC
     451  NGGGATATGG GCAGGTGCT GGAACATTAC GCCGCGAGCG GTTTTGCCCA
20      501  GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
     551  CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC
     601  CAACAGGCGG GTTCGGTCAG GGATNNGGAA AGCATAGGCG GCGTATTGTA
     651  CGCGCANGGC TGGCTGTCGG CAGNNACGCA CAACGGGCGC GATTACGCCT
     701  TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
25      751  ATCGAAAAGG CAAGGGCGNA ANANNNTNAG TTGAGTTACA GCAAAAAGG
     801  TTTGCAGACC TTTTTCCTNG CAACCCTGCT GATTGCTCN CTGCTGTCGA
     851  TTTTCTTGTC ACTGGTCATG GCACTGTATT TCGCCCGCCG TTTCTGCGAA
     901  CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT
     951  CAGCCAGACG CGCCCCGTGT TCGCACAAGA CGAGTTCGGA CGCTTGACCA
30     1001  AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCAA AGAAGCAGAC
     1051  GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
     1101  GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
     1151  TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
     1201  CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
35     1251  CTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGCGC GCAGGTACGG
     1301  ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
     1351  CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACNGCAACG GCGTGGAAT
     1401  GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAA GAAGCCGCGT
     1451  GGGGCGAAGT GGCAAAACCG CTGGCACACG AAATCCGCAA TCCGCTCACG
40     1501  CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
     1551  GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
     1601  AACAAGTGGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
     1651  CGTTCCCTT CGNCTCAATT GGAAAATCAG GATTTGAACG CCTTAATCGG
     1701  CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGTTT GCGGCGGAAC
45     1751  TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
     1801  GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
     1851  TGTGCCCCGAA GTCAGGGTAA AATCGGAAGC GGGGCGAGAC GGACGGATTG
     1901  TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC
     1951  AATGACCTTC AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
50     2001  ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
     2051  TGAGCAATCA GGATGCGGGC GCGCGTNTG TCAGAATCAT CTTGCCAAAA
     2101  ACGGTAGAAA CTTATGCGTA G

```

This encodes a protein having amino acid sequence (SEQ ID NO: 254):

```

55      1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
     51  LARYVILLLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING

```

5  
10  
15  
20  
25  
30  
35  
40  
45  
50

```

101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDxE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAAQSSL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551 RSPSXQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLM MAADTTAMRQ
601 VLHNIFKNAA EAAEEADVPE VRVKSEAGQD GRIVLTVCDN GKGFGREMLH
651 NAFEPYVTDK PAGTGLXLPV VKKIIEEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*

```

15 ORF64a (SEQ ID NO: 254) and ORF64-1 (SEQ ID NO: 252) show 96.6% identity in 706 aa overlap:

20  
25  
30  
35  
40  
45  
50

```

              10      20      30      40      50      60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAMLLLVLSAVLARYVILLK
              |||||
orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAMLLLVLSAVLARYVILLK
              10      20      30      40      50      60

              70      80      90     100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              |||||
orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              70      80      90     100     110     120

              130     140     150     160     170     180
orf64a.pep  SKSALNLAADNALGNAIPVQIDIXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              |||||:|||||
orf64-1      SKSALNLAADNALGNAIPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
              130     140     150     160     170     180

              190     200     210     220     230     240
orf64a.pep  SINPHKLDQPPFPGKARWEKIQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQPV
              |||||:|||||
orf64-1      SINPHKLDQPPFPGKARWEKIQAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV
              190     200     210     220     230     240

              250     260     270     280     290     300
orf64a.pep  PKGVAEDAVLIEKARAXXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
              |||||
orf64-1      PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
              250     260     270     280     290     300

              310     320     330     340     350     360
orf64a.pep  PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA
              |||||
orf64-1      PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA
              310     320     330     340     350     360

              370     380     390     400     410     420
orf64a.pep  RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAAQSSL
              |||||
orf64-1      RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAAQSSL
              370     380     390     400     410     420

```

		430	440	450	460	470	480
	orf64a.pep	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMIDDITVLIHAQK					
5	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMIDDITVLIHAQK					
		430	440	450	460	470	480
	orf64a.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDTIIKQVAALK					
10	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDTIVKQVAALK					
		490	500	510	520	530	540
	orf64a.pep	EMVEAFRNYXRSPSXQLENQDLNALIGDVLALYEAGPCRFAAELAGEPLMMAADTTAMRQ					
15	orf64-1	EMVEAFRNYARSPSLKENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ					
		550	560	570	580	590	600
	orf64a.pep	VLHNI FKNAAEAAEEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK					
20	orf64-1	VLHNI FKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK					
		610	620	630	640	650	660
	orf64a.pep	PAGTGLXLPVVKKII EEHGGXISLSNQDAGGAXVRI ILPKTVETYAX					
25	orf64-1	PAGTGLGLPVVKKII EEHGGRI SLSNQDAGGACVRI ILPKTVKTYAX					
		670	680	690	700		

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF64 (SEQ ID NO: 250) shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) (SEQ ID NO: 256) from *N. gonorrhoeae*:

30	orf64.pep	MRRFLPIAAICAXLXXGLTAATGSTSSLADYFWWIVAFSAMPLLVLSAVLARYVILLK	60
	orf64ng	MRRFLPIAAICAVLLYGLTAATGSTSSLADYFWWIVSFSAMPLLVLSAVLARYVILLK	60
	orf64.pep	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
35	orf64.pep	LSKSALNLAADNALGNAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng	LSKSALDLAADNAVSNAPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIE	179
	orf64.pep	KSINPHKLDQPPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
40	orf64ng	KSINPHQFDQPLPDKEHWEIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239
	orf64.pep	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	300
	orf64ng	IPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLLIASLLSIFLALVMALYFARRFV	299

```

orf64 .pep      EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTXLFNHMTEQLSIAKDADERNRRREEA  360
                ||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||:
orf64ng         EPILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEA  359

orf64 .pep      ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT                394
                |||||||:|||||||:|:|
orf64ng         ARHYLECVLDGLTTGVVVSYP LSCCRTAVFSTCHSSPLSYF  400

```

An ORF64ng nucleotide sequence (SEQ ID NO: 255) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 256):

```

10      1  MRRFLPIAAI  CAVVLLYGLT  AATGSTSSLA  DYFWWIVSFS  AMLLLVLSAV
        51  LARYVILLLK  DRRNGVFGSQ  IAKRLSGMFT  LVAVLPGLFL  FGISAQFING
       101  TINSWFGNDT  HEALERSLNL  SKSALDLAAD  NAVSNAVVPQ  IDLIGTASLS
       151  GNMGSVLEHY  AGSGFAQLAL  YNAASGKIEK  SINPHQFDQP  LPDKEHWEQI
       201  QQTGSVRSLE  SIGGVLYAQQ  WLSAGTHNGR  DYALFFRQPI  PENVAQDAVL
15      251  IEKARAKYAE  LSYSKKGLOT  FFLVTLIIAS  LLSIFLALVM  ALYFARRFVE
       301  PILSLAEGAK  AVAQGDFSQT  RPVLNRNDEFG  RLTKLFNHMT  EQLSIAKEAD
       351  ERNRRREEAA  RHYLECVLDG  LTTGVVVSYP  LSCCRTAVFS  TCHSSPLSYF*

```

Further work revealed the complete gonococcal DNA sequence (SEQ ID NO: 257):

```

20      1  ATGCGCCGCT  TCCTACCGAT  CGCAGCCATA  TCGCGCCGTCG  TCCTGCTGTA
        51  CGGATTGACG  GCGGCGACCG  GCAGCACCAG  TTCGCTGGCG  GATTATTTCT
       101  GGTGGATAGT  CTCGTTCCAGC  GCAATGCTGC  TGCTGGTGTT  GTCCGCCGTT
       151  TTGGCACGTT  ATGTCATATT  GCTGTTGAAA  GACAGGCGCA  ACGGCGTGTT
       201  CGGTTCCGAG  ATTGCCAAAC  GCCTTTCCGG  GATGTTACAG  CTGGTCGCCG
25      251  TACTGCCCGG  CTTGTTCCCTG  TTCGGCATT  CCGCGCAGTT  TATCAACGGC
       301  ACGATTAATT  CGTGGTTCGG  CAACGACACC  CACGAAGCCC  TCGAACGCAG
       351  CCTTAATTTG  AGCAAGTCCG  CACTGGATT  GGCGGCAGAC  AATGCCGTCA
       401  GCAACGCCGT  TCCCGTACAG  ATAGACCTCA  TCGGCACCGC  CTCCCTGTCT
       451  GGCAATATGG  GCAGTGTGCT  GGAACACTAC  GCCGGCAGCG  GTTTTGCCCA
30      501  GCTTGCCCTG  TACAATGCCG  CAAGCGGGAA  AATCGAAAAA  AGCATCAATC
       551  CGCACCAATT  CGACCAGCCG  CTTCCCGACA  AAGAACATTG  GGAACAGATT
       601  CAGCAGACCG  GTTCGGTTCG  GAGTTTGGA  AGCATAGGCG  GCGTATTGTA
       651  CGCGCAGGGA  TGGTTGTCGG  CAGGTACGCA  CAACGGGCGC  GATTACGCGC
       701  TGTTCTTCCG  CCAGCCGATT  CCCGAAAATG  TGGCACAGGA  TGCCGTCTCT
35      751  ATTGAAAAGG  CGCGGGCGAA  ATATGCCGAA  TTGAGTTACA  GCAAAAAAGG
       801  TTTGCAGACC  TTTTTTCTGG  TAACCCTGCT  GATTGCCTCG  CTGCTGTCTG
       851  TTTTCTTCTG  GCTGGTAATG  GCACTGTATT  TTGCCCGCCG  TTTCTGTCGA
       901  CCCATTCTGT  CGCTTGCCGA  GGGCGCAAAG  GCGGTGGCGC  AGGGTGATTT
       951  CAGCCAGACG  CGCCCCGTAT  TGCGCAACGA  CGAGTTCGGA  CGTTTGACCA
40     1001  AGCTGTTCAA  CCATATGACC  GAGCAGCTTT  CCATCGCCAA  AGAAGCAGAC
       1051  GAACGCAACC  GCCGGCGCGA  GGAAGCCGCC  CGTCACTACC  TCGAGTGCCT
       1101  GTTGGATGGG  TTGACTACCG  GTGTGGTGGT  GTTTGACGAA  AAAGGCCGTT
       1151  TGAAAACCTT  CAACAAGGCG  GCGGAACAGA  TTTTGGGGAT  GCCGCTCGCC
       1201  CCCCTGTGGG  GCAGCAGCCG  GCACGGTTGG  CACGGCGTTT  CGGCGCAGCA
45     1251  GTCCCTGCTT  GCCGAAGTGT  TtgccgccAT  CCGTGCGGCG  GCAGGTACGG
       1301  ACAAACCGGT  CCAGGTGGAA  TATGCCGCGC  CGGACGATGC  CAAAATCCTG
       1351  CTGGGCAAGG  CGACGGTATT  GCCCGAAGAC  AACGGCAACG  GCGTGGTGAT
       1401  GGTGATTGAC  GACATCACCG  TGCTGATACG  CGCGCAAAAA  GAAGCCGCGT
       1451  GGGGTGAAGT  GGCGAAGCGG  CTGGCACACG  AAATCCGCAA  TCCGCTCACG
50     1501  CCCATCCAGC  TTTCCGCCGA  ACGGCTGGCG  TGGAAATTGG  GCGGGAAGCT
       1551  GGACGATCAG  GACGCGCAAA  TCCTGACGCG  TtcgACCGAC  ACCATCATCA
       1601  AACAGgtggc  gGCGTTAAAA  GAAATGGTCG  AGGCATTCCG  CAATTACGCG
       1651  CGCGCCCCCT  CGCTCAAAC  GGAAAATCAG  GATTTGAACG  CCTTAATCGG
       1701  CGATGTTTTG  GCCCTGTACG  AAGCCGGCCC  GTGCCGGTTT  GAGGCGGAAC

```



-235-

1751 TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCCGGCAG  
 1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA  
 1851 TATGCCCGAA GTCAGGGTAA AATCGGAAAC GGGGCAGGAC GGACGGATTG  
 1901 TCCTGACGGT TTGCACAAC GGCAAGGGAT TCGGCAAGGA AATGCTGCAC  
 1951 AATGCTTTCG AGCCGTATGT GACGGATAAG CCGGCGGGAA CGGGACTGGG  
 2001 TCTGCCTGTA GTGAAAAAA TCATTGGAGA ACACGGCGGC CGCATCAGCC  
 2051 TGAGCAATCA GGATGCGGGT GGGGCGTGTG TCAGAATCAT CTTGCCAAAA  
 2101 ACGGTAGAAA CTTATGCGTA G

10 This corresponds to the amino acid sequence (SEQ ID NO: 258; ORF64ng-1):

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV  
 51 LARYVILLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING  
 101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVVPQ IDLIGTASLS  
 151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQ LPDKEHWEQI  
 15 QQTGVSRSLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPI PENVAQDAVL  
 201 IEKARAKYAE LSYSKKGLQT FFLVTLIIAS LLSIFLALVM ALYFARRFVE  
 251 PILSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD  
 301 ERNRRREEAA RHYLCVLDG LTTGVVVFDE KGRLKTFNKA AEQILGMPLA  
 351 PLWGSRRHWG HGVSAQSSLL AEVFAAIGAA AGTDKPVQVE YAAPDDAKIL  
 401 LGKATVLPED NGNGVVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT  
 451 PIQLSAERLA WKLGGKDDQ DAQILTRSTD TIIKQVAALK EMVEAFRNYA  
 501 RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAELAGEPLM MAADTTAMRQ  
 551 VLHNIPKNAA EAAEEADMPE VRVKSETGQD GRIVLTVCDN GKGFGEMLH  
 601 NAFEPYVTDK PAGTGLGLPV VKKIIGEHGG RISLSNQDAG GACVRIILPK  
 651 TVETYA\*

ORF64ng-1 (SEQ ID NO: 258) and ORF64-1 (SEQ ID NO: 252) show 93.8% identity in 706 aa overlap:

30	orf64ng-1.pep	MRRFLPIAAICAVVLLYGLTAAATGSTSSLDYFWWIVSFSAMLLLVLSAVLARYVILLK
	orf64-1	MRRFLPIAAICAVVLLYGLTAAATGSTSSLDYFWWIVAFSAMLLLVLSAVLARYVILLK
35	orf64ng-1.pep	DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
	orf64-1	DRRDGVFGSQIAKRLSGMFTLVAVLPGLVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
40	orf64ng-1.pep	SKSALDLAADNAVSNVAVPQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
	orf64-1	SKSALNLAADNALGNVAVPQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
45	orf64ng-1.pep	SINPHQFDQPLPDKEHWEQIQQTGVSRSLESIGGVLYAQQWLSAGTHNGRDYALFFRQPI
	orf64-1	SINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQQWLSAGTHNGRDYALFFRQPV
50	orf64ng-1.pep	PENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLIIASLLSIFLALVMALYFARRFVE

-236-

5	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE	250	260	270	280	290	300
	orf64ng-1.pep	PILSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTKLFNHMTQLSIAKEADERNNRREEAA	310	320	330	340	350	360
10	orf64-1	PVLSLAEGAKAVAQGDIFSQTRPVLNRNDEFGRLTKLFNHMTQLSIAKEADERNNRREEAA	310	320	330	340	350	360
	orf64ng-1.pep	RHYLECVLDGLTTGVVVVFDEKGRLLKTFNKAAEQILGMPLAPLWGSSRHGWHGVSAQQSLL	370	380	390	400	410	420
15	orf64-1	RHYLECVLEGLTTGVVVVFDEQGLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL	370	380	390	400	410	420
	orf64ng-1.pep	AEVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIQAQK	430	440	450	460	470	480
20	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK	430	440	450	460	470	480
	orf64ng-1.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDITIKQVAALK	490	500	510	520	530	540
25	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTDITIVKQVAALK	490	500	510	520	530	540
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLMMAADTTAMRQ	550	560	570	580	590	600
30	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ	550	560	570	580	590	600
	orf64ng-1.pep	VLHNIFKNAAEAAEEADMPVVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
35	orf64-1	VLHNIFKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK	610	620	630	640	650	660
	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVETYAX	670	680	690	700		
40	orf64-1	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVKTYAX	670	680	690	700		

Furthermore, ORF64ng-1 (SEQ ID NO: 258) shows significant homology to a protein (SEQ ID

NO: 1129) from *A.caulinodans*:

sp|Q04850|NTRY\_AZOCA NITROGEN REGULATION PROTEIN NTRY )gi|77479|pir||S18624 ntry  
protein - Azorhizobium caulinodans )gi|38737 (X63841) NtrY gene product  
[Azorhizobium caulinodans] Length = 771  
Score = 218 bits (550), Expect = 7e-56  
Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)

Query: 7 IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXXRYVILLKDRRNGV 66  
I+A+ ++L GLT + + + R + + K R G  
Sbjct: 35 ISALATFLILMGLTPVVPVTHQVVIS----VLLVNAAAVLILSAMVGREIWRIAKARAGR 90

Query: 67 FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD 126  
 +++ R+ G+F +V+V+P + + +++ ++ ++ WF T E + S+++++ +  
 Sbjct: 91 AAARLHIRIVGLFAVVSVPAILVAVVASLTLDRLDRWFSMRTQEIVASSVSVAQTYVR 150

5 Query: 127 LAADNAVSNAPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184  
 A N + + + DL S+ Y G S F Q+ AA + ++  
 Sbjct: 151 EHALNIRGDILAMSADLTRLKSV-----YEGDRSRFNQILTAQAALRNLPGAMLI 200

Query: 185 HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWSAGTHNGRDYA----- 233  
 + D + ++ + I + V + +IG Q + N DY  
 Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQPVIYLP--NDADYVAAVVPLKDYDD 256

10 Query: 234 --LFFRQPIPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTXXXXXXXXXXXXXVMA 291  
 L+ + I V ++ A Y L + G+Q F + +  
 Sbjct: 257 LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWLG 316

Query: 292 LYFARRFVEPILSLAEGAKAVAQGDFSQTRPVLNRD-EFGRLTKLFNHMTQELSIXXXXX 350  
 L F++ V PI L A VA+G+ P+ R + + L + FN MT +L  
 15 Sbjct: 317 LNFQSKWLVAPIRRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI 376

Query: 351 XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRKTFNKAAEQILGMPLAPLWGSSRHGW 410  
 + E VL G+ GV+ D + R+ N++AE++LG L+ + RH  
 Sbjct: 377 LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSARLLG--LSEVEALHRHLA 434

20 Query: 411 HGVSAQQSLLAEVFXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM 467  
 V LL E + VQ D + + V E + +G V+  
 Sbjct: 435 EVVPETAGLLEEA-----EHARQRSVQGNITLTRDGRERVFAVRVTTEQSPEAEHGWV 488

Query: 468 VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLLDDQDAQILTR 527  
 +DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +  
 Sbjct: 489 TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ 547

25 Query: 528 STDIIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE 587  
 TDTII+QV + MV+ F ++AR P +++QD++ +I + L G +  
 Sbjct: 548 CDTIIRQVGDIGRMVDEFSSFARMKPKVVDSDQMSEIIRQTVFLMRVGHPEVVFDSSEVP 607

Query: 588 PLMAA-DTTAMRQVLHNIFKNXXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD 639  
 P M A D + Q L NI KN P+VR + + G+D +V+ + D  
 30 Sbjct: 608 PAMPARFDRRLVSQALTNILKNAAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664

Query: 640 NGKGFGKEMLHNAFEPYVTDKPGTGLGLPVVKKIIGEHHGRISLSNQDAG-GACVRIIL 698  
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L  
 Sbjct: 665 NGTGLPQESRNRLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined)  
 35 and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is  
 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
 useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 31

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 259):

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT  
 51 GCTTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC  
 101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC  
 151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT  
 5 201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT  
 251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG  
 301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC  
 351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC  
 401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG  
 10 451 CACGCGTTGG ATACG...

This corresponds to the amino acid sequence (SEQ ID NO: 260; ORF66):

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQPPFQIFGI HTTPGAFSFP  
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA  
 15 101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPNASTVIG  
 151 HALDT...

Further work revealed the complete nucleotide sequence (SEQ ID NO: 261):

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT  
 51 GCTTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC  
 20 101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC  
 151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT  
 201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT  
 251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG  
 301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC  
 25 351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC  
 401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC  
 451 AACGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG  
 501 CGATGGATT TATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC  
 551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCCG CTACGGCGTG  
 30 601 ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC  
 651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

This corresponds to the amino acid sequence (SEQ ID NO: 262; ORF66-1):

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQPPFQIFGI HTTPGAFSFP  
 35 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA  
 101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPTASTVIG  
 151 NALDTLVFFA VAFYASSDGF MAANWQGIAP VDYLFLKLTVC TLFFLPAYGV  
 201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP\*

40 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 (SEQ ID NO: 1130) of *E. coli* (accession number P37619)

ORF66 (SEQ ID NO: 260) and o221 protein (SEQ ID NO: 1130) show 67% aa identity in 155aa overlap:

```

orf66  1  MYAFTAAQQQKALFRLVLFHILIIAASNVLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
        M  F+  Q+  KALF L LFH+L+I +SNYLVQ P  I G HTTWGAFSFPFIFLATDLTV
o221   1  MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

orf66  61  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
        RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
o221   61  RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

orf66  121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
        +GQILD+ VFN+LR+ + WW+AP AST+ G+  DT
o221   121 LGQILDVHVFNRLRQSRRWWLAPTASTLFGNVSDT 155

```

# 10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 (SEQ ID NO: 260) shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) (SEQ ID NO: 264) from strain A of *N. meningitidis*:

```

15 orf66.pep  10      20      30      40      50      60
    MYAFTAAQQQKALFRLVLFHILIIAASNVLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf66a      10      20      30      40      50      60
    MYAFTAAQQQKALFWLVLFHILIIAASNVLVQFPFQISGIHTTWGAFSFPFIFLATDLTV

20 orf66.pep  70      80      90      100     110     120
    RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf66a      70      80      90      100     110     120
    RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA

25 orf66.pep  130     140     150
    IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT
    :|||||||||||||||||:|:|:|:|:|:|:|
orf66a      130     140     150     160     170     180
    LGQILDIFVFNKLRLKAWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAP

orf66a      190     200     210     220
    VDYLFKLTVCGFLFLPAYGVILNLLTKKLTTLTQTKAQDRPAPSLQNPX

```

The complete length ORF66a nucleotide sequence (SEQ ID NO: 263) is:

```

35 1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
    51 GCTTTTTCAT ATCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
    101 CCTTCCAAAT TTCCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
    151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
    201 GGCACGGCGG ATTATCTTTT GGGTCATGTT CCCC GCCCTT TTGCTTTCCT
    251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGACGGG CTGGGCGCG
    301 CTGTCCGAAT TCAACACCTT TGTCCGACGC ATCGCGCTGG CAAGTTTTCG
    351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTC AACAAATTAC
    401 GCCGTCTGAA AGCGTGGTGG GTTGCCCGCA CTGCATCAAC CGTCATCGGC
    451 AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCTTCT ACGCAAGCAG
    501 CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
    551 TGTTCAAAC TACCGTCTGC GGTCTGTTT TCCTGCCCGC CTACGGCGTG
    601 ATTCTGAATC TGCTGACGAA AAAACTGACG ACCCTGCAAA CCAAACAGGC
    651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 264):

1 MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP  
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA  
 101 LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRRKAWW VAPTASTVIG  
 5 151 NALDTLVFFA VAFYASSDGF MAANWQGI AF VDYLFKLTVC GLFFLPAYGV  
 201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP\*

ORF66a (SEQ ID NO: 264) and ORF66-1 (SEQ ID NO: 262) show 97.8% identity in 228 aa overlap:

10		10	20	30	40	50	60
	orf66a.pep	MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV					
	orf66-1	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV					
		10	20	30	40	50	60
15		70	80	90	100	110	120
	orf66a.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
	orf66-1	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf66a.pep	LGQILDIFVFNKLRRKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
	orf66-1	IGQILDIFVFNKLRRKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
		130	140	150	160	170	180
25		190	200	210	220	229	
	orf66a.pep	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
	orf66-1	VDYLFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
		190	200	210	220		

### 30 Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 (SEQ ID NO: 260) shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) (SEQ ID NO: 266) from *N. gonorrhoeae*:

35	orf66.pep	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66ng	MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
40	orf66.pep	IGQILDIFVFNKLRRKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRRKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF	180

The complete length ORF66ng nucleotide sequence (SEQ ID NO: 265) is:

1 ATGTACGCAT TGACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT  
 51 GCTTTTCCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC  
 101 CCTTCCGGAT TTTCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC  
 151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT  
 5 201 GGC GCGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT ttgCTTTcat  
 251 aCGTCTTTTC CGTTTGTTC CACAACGGCA GTTGGACGGG CTGGGCGCG  
 301 ctgTCCCAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTGC  
 351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTCGTATTC GACAAATTAC  
 401 GCCGTCTGAA AGCGTGTTGG ATTGCCCCGG CCGCATCAAC CGTCATCGGC  
 10 451 AATGCACTGG ACACGTTAGT ATTTTGTGCC GTTGCCTTTT ACGCAAGCAG  
 501 CGATGAATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC  
 551 TGTTCAAACT TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG  
 601 ATACTGAATC TGCTGACGAA AAAACTGACG GCCTGCAAA CCAAACAGGC  
 15 651 GCAAGACCGC CCCGTGCCCT CGCTGCAAAA TCCGTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 266):

1 MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQPPFRIFGI HTTWGAFSFP  
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL SLSYVFSVLF HNGSWTGLGA  
 101 PSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRLKAWW IAPAASTVIG  
 20 151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYL FKLTV C TLFFLPAYGV  
 201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP\*

An alternative annotated sequence is:

1 MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQPPFRIFGI HTTWGAFSFP  
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA  
 101 LSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRLKAWW IAPAASTVIG  
 151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYL FKLTV C TLFFLPAYGV  
 201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP\*

30 ORF66ng (SEQ ID NO: 266) and ORF66-1 (SEQ ID NO: 262) show 96.1% identity in 228 aa overlap:

orf66-1.pep MYAFTAAQQQKALFRLVLFHILIIAASNYLVQPPFQIFGIHTTWGAFSFPFIFLATDLTV 60  
 orf66ng MYALTAAQQQKALFRLVLFHILIIAASNYLVQPPFRIFGIHTTWGAFSFPFIFLATDLTV 60  
 35 orf66-1.pep RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGR IALASFAAYA 120  
 orf66ng RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGR IALASFAAYA 120  
 orf66-1.pep IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180  
 40 orf66ng LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180  
 orf66-1.pep VDYL FKLTV C TLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX 229  
 orf66ng VDYL FKLTV C TLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229

45 Furthermore, ORF66ng (SEQ ID NO: 266) shows significant homology with an *E.coli* ORF (SEQ ID NO: 1130):

```

sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC REGION
(O221)
)gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli )gi|466607
(U00039) No definition line found [Escherichia coli] )gi|1789882 (AE000423)
5 hypothetical 25.3 kD protein in ftsY-nika intergenic region [Escherichia coli]
Length = 221
Score = 273 bits (692), Expect = 5e-73
Identities = 132/203 (65%), Positives = 155/203 (76%)

Query: 1 MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV 60
10 M + Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
Sbjct: 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

Query: 61 RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
Sbjct: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASFMAYA 120

15 Query: 121 LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAP 180
LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
Sbjct: 121 LGQILDVHVFNRRLQSRWWLAPTASTLFGNVSDTLAFFFIAPFWRSPDAFMAEHWMEIAL 180

Query: 181 VDYLFKLTVCTLFFLPAYGVILN 203
VDY FK+ + +FFLP YGV+LN
20 Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203

```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 25 Example 32

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 267):

```

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAyGCA GTmwrAATAT
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
30 151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201 TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC
301 CGCTTAGgCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
351 CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
35 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAAA AGGCTACGAA
451 TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA
501 TGGCTGCTAC GCGGTTGAT.

```

This corresponds to the amino acid sequence (SEQ ID NO: 268; ORF72):

```

40 1 MVIKYTNLNF AKLSIIILM MYSFEANANA VXISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKEQ IQARGYQYDP ETDKFKVGYE
151 YSNCLWYEDK RRINRTYGCY GVD..

```



Further work revealed the complete nucleotide sequence (SEQ ID NO: 269):

```

1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
5  51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTAAATAA AAACGGTAGA
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
10 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351 CCTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC
451 TAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 270; ORF72-1):

```

15 1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51  VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF72 (SEQ ID NO: 268) shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) (SEQ ID NO: 272) from strain A of *N. meningitidis*:

```

25 orf72.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VXiSETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
orf72a      MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
           10      20      30      40      50      60

30 orf72.pep      70      80      90      100     110     120
    DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA
    |||||
orf72a      DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA
    |||||
           70      80      90      100     110     120

35 orf72.pep      130     140     150     160     170
    HDVYETFKED IQARGYQYDP ETDKFVKGYEYSNCLWYED KRRINR TYGCGYGV
    |||||
orf72a      HDVYETFKED IQARGYQYDP ETDKFAKVS GX
    |||||
           130     140     150

```

40 The complete length ORF72a nucleotide sequence (SEQ ID NO: 271) is:

```

1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTAAATAA AAACGGTAGA

```



15 After further analysis, the following gonococcal DNA sequence (SEQ ID NO: 275) was identified:

This corresponds to the amino acid sequence (SEQ ID NO: 276; ORF72ng-1):

ORF72ng-1 (SEQ ID NO: 276) and ORF721-1 (SEQ ID NO: 270) show 89.7% identity in 145 aa overlap:

		10	20	30	40	50	60
35	orf72ng-1.pe	MVTKHTNLNFAKL	SIIAILMMYSFEANANAVK	ISETLSVDTGGQAKVHKFVPKSSNIYSS			
		:			:	:	:
	orf72-1	MVIKYTNLNFAKL	SIIAILMMYSFEANANAVK	ISETVSVDTGGQAKIHKFVPKN	SKTYSS		
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf72ng-1.pe	DLTKAVDLTHIPTGAKARINAKITASVSRAGVL	SGVGKLV	RQGAKFGTRAPV	PYVGTTALLA		
		:			:	:	
	orf72-1	DLIKTVDLTHIPTGAKARINAKITASVSRAGVL	AGVGKLR	LARGAKFPSTRAPV	PYVGTTALLA		
		70	80	90	100	110	120
		130	140				
45	orf72ng-1.pe	HDVYETFKEDIQARGCRYDPETDKF					
				:			
	orf72-1	HDVYETFKEDIQARGYQYDPETDKFAKVSGX					
		130	140	150			

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 5 Example 33

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 277):

```

1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
151 GCTGACCGGT CTTTATTGGG CGGGCGCGGC AATGAGAAGC GCGGGAAGG
201 TATCCGTTTA TCAGATGTTG TGGCCTATC..

```

This corresponds to the amino acid sequence (SEQ ID NO: 278; ORF73):

```

1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLQTG
151 LTGLLLAGAA MRSGGKVSIV QMLWPI..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 279):

```

1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101 TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCATACGGGG
151 CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201 ATCCGTTTAT CAGATGTTGT GGCCATCCG TTATACGGTG GCGGCTGTGT
251 GTCTGATGAG TCCGGGATTC GTATCCTCGG TGTGCGCGGT ATTGCTGCTG
301 CTGCCGTTTA AGGGAGGGGC AGTGTTCAG GCAGGAGGTG CGGAAAATTT
25  351 TTTCAACATG AACCAATCGG GCAGAAAAGA GGGCTTTTCC CGCGATGACG
401 ATATTATCGA GGGAGAAATAT ACGGTTGAAG AGCCTTACGG CGGCAATCGT
451 TCCCGAAACG CCATCGAACA CAAAAAAGAC GAATAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 280; ORF73-1):

```

30  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRHTG
51  LSGLLLAGAA MRSGGRVSIV QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEPYGGNR
151 SRNAIEHKKD E*

```

35 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 (SEQ ID NO: 278) shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) (SEQ ID NO: 282) from strain A of *N. meningitidis*:

5  
10

orf73.pep	10	20	30	40	50	60
	MRFFGIGFLVLLFLEIMSI	VWVADWLGGGWT	FLMAAGFAAGVLM	LRQTGLTGLLLAGAA		
orf73a	MRFFGIGFLVLLFLEIMSI	VWVADWLGGGWT	FLMAATFAAGVVML	RHTGLSGLLLAGAA		
	10	20	30	40	50	60
orf73.pep	70					
	MRSGGKVS	VYQMLWPI				
orf73a	MRSGGRVSVYXMLWXIRY	TVAAVCX	MSPGFVSSVXAVLLXL	LPFKGGAVLQAGGAENFFNM		

The complete length ORF73a nucleotide sequence (SEQ ID NO: 281) is:

15  
20

1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
51	GTCGATTGTG	TGGGTTGCCG	ATTGGTTGGG	CGGCGGTTGG	ACGCTGTTTC
101	TAATGGCGGC	AACCTTTGCC	GCCGGCGTGG	TGATGCTCAG	GCATACGGGG
151	CTGTCCGTC	TTTTATTGGC	GGGCGCGGCA	ATGAGAAGCG	GCGGGAGGGT
201	ATCCGTTTAT	CANATGTTGT	GGCNTATCCG	TTATACGGTG	GCGGCGGTGT
251	GTCNGATGAG	TCCGGGATTC	GTATCCTCGG	TGTNGGCGGT	ATTGCTGNTG
301	CTNCCGTTTA	AGGGAGGTGC	AGTGTTCAG	GCAGGAGGTG	CGGAAAATTT
351	TTTCAACATG	AACCANTCGG	GCAGAAAAGA	NGGCNTTTC	CGCGATGACG
401	ATATTATCGA	GGGGGAATAT	ACGGTTGAAG	ANCCTTACGG	CGGCANTCGT
451	TTCCGAAACG	CCNTNGAACA	CAAAAAGAC	GAATAA	

This encodes a protein having amino acid sequence (SEQ ID NO: 282):

25

1	MRFFGIGFLV	LLFLEIMSI	VWVADWLGGGW	TLFLMAATFA	AGVVMRLRHTG
51	LSGLLLAGAA	MRSGGRVSVY	XMLWXIRYTV	AAVCXMSPGF	VSSVXAVLLX
101	LPFKGGAVLQ	AGGAENFFNM	NXSGRKXGXS	RDDDIIEGEY	TVEXPYGGXR
151	FRNAXEHKKD	E*			

30  
ORF73a (SEQ ID NO: 282) and ORF73-1 (SEQ ID NO: 280) show 91.3% identity in 161 aa overlap

35  
40  
45

orf73a.pep	10	20	30	40	50	60
	MRFFGIGFLVLLFLEIMSI	VWVADWLGGGWT	FLMAATFAAGVVML	RHTGLSGLLLAGAA		
orf73-1	MRFFGIGFLVLLFLEIMSI	VWVADWLGGGWT	FLMAAGFAAGVLM	RHTGLSGLLLAGAA		
	10	20	30	40	50	60
orf73a.pep	70	80	90	100	110	120
	MRSGGRVSVYXMLWXIRY	TVAAVCX	MSPGFVSSVXAVLLXL	LPFKGGAVLQAGGAENFFNM		
orf73-1	MRSGGRVSVYQMLWPIRY	TVAAVCL	MSPGFVSSVLAVLLLL	LPFKGGAVLQAGGAENFFNM		
	70	80	90	100	110	120
orf73a.pep	130	140	150	160		
	NXSGRKXGXS	RDDDIIEGEY	TVEXPYGGXR	FRNAXEHKKDEX		
orf73-1	NQSGRKEGFS	RDDDIIEGEY	TVPEPYGGNRS	RNAIEHKKDEX		
	130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 (SEQ ID NO: 278) shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) (SEQ ID NO: 284) from *N. gonorrhoeae*:

```

5  orf73.pep      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLQTGLTGLLLAGAA      60
    |||||||||||||||||||||||||||||||||||||||||||||||||:|:|:|
    orf73ng       MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVLMRLHTGLSGLLLAGAA      60

    orf73.pep      MRSGGKVSQMLWPI                                          76
    ::|||
    orf73ng       VKSSGKVSQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM  120

```

10 The complete length ORF73ng nucleotide sequence (SEQ ID NO: 283) is:

```

1  ATGAGATTTT TCGGTATCGG TTTTGGTG CTGCTGTTT TGGAAATTAT
51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG AcgcTGTTTC
101 TAATGGCGGC AACCTTTGCC GCCGTGTGC TGATGCTCAG GCATAcggGG
151 CTGTCCGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
15  201 ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
    251 GTCTGatgag tCcgGATTC GTATCCTccg tgttggCGGT ATTGCTGCTG
    301 CTGCCgttta aggGaggGgc agtgttcag gcaggagggtg cggaaaATTT
    351 TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
    401 atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
20  451 tcccgAAAcg ccatcgaaca cgaaaAagac gaataA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 284):

```

1  MRFFGIGFLV LLFLEIMSIW VVADWLGGGW TLFLMAATFA AGVLMRLHTG
25 51  LSGLLLAGAA VKSSGKVSQY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101 LPFGKGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGRN
151 SRNAIEHEKD E*

```

ORF73ng (SEQ ID NO: 284) and ORG73-1 (SEQ ID NO: 280) show 93.8% identity in 161 aa overlap

```

30          10      20      30      40      50      60
    orf73-1.pep  MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLHTGLSGLLLAGAA
    orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVLMRLHTGLSGLLLAGAA
          10      20      30      40      50      60

35          70      80      90      100     110     120
    orf73-1.pep  MRSGGRVSQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM
    orf73ng      VKSSGKVSQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM
          70      80      90      100     110     120

40          130     140     150     160
    orf73-1.pep  NQSGRKEGFSRDDDIIEGEYTVVEPYGGRNRSRNAIEHKKDEX
    orf73ng      NQSGRKEGFFHDDDIIEGEYTVVEKPDGGRNRSRNAIEHEKDEX
          130     140     150     160

45

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 34

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 285):

```

1  ATGTTTGT TTTTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
101 TCGGCAATTT GCGCGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
151 GCG..... GCGGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
201 CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
251 GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
301 GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
351 ACTCGCCCGC CGCGTGCCTG AGGCCGGGTT TAAAGTCGTT CCCGTGCTGG
15 401 GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
451 GATTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA
501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
551 CGCCGACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
20 651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
701 ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
751 AAACACGAAG GCTTGTCGGA GTCCGCGCAA AACATCATGA AAATCCTCAC
801 AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG
25 851 GCGAGGGAAA GAAAGCTTTG TACGAT..

```

This corresponds to the amino acid sequence (SEQ ID NO: 286; ORF75):

```

1  MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
51  A....AEDTR VTAQLLSAYG IQKLVSVRE HNERQMADKI VGYLSDGMVV
101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVY PVVGAXAVMA ALSVAGVEGS
30 151 DFYFNGFVPP KSGERRKLFA KVVRAAFPIV MFETPHRIGA ALADMAELFP
201 ERLMLAREI TKTFETPLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE
251 KHEGLSESAQ NIMKILTAEL PTKQAAELAA KITGEGKKAL YD..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 287):

```

35 1  ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGCGTTACCG CACAGCTTTT GAGCGGTAC GGCATTACAG GCAAACCTCGT
201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGCT
40 251 ATCTTTTACA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTTCAACGG TTTGTACCG
451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGC
45 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
601 ATTACGAAAA CGTTTGAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA

```

5  
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTGG  
 701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG  
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
 801 GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC  
 851 TGGCTCTGTC TTGGAATAAC AAATAG

This corresponds to the amino acid sequence (SEQ ID NO: 288; ORF75-1):

10  
 1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQKGLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE  
 201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

15 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 (SEQ ID NO: 286) shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) (SEQ ID NO: 290) from strain A of *N. meningitidis*:

20  
 orf75.pep MFVFQTAFXMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKAXXXAEDTR  
 orf75a MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTR  
 10 20 30 40 50 60

25  
 orf75.pep VTAQLLSAYGIQKGLVSVREHNERQMA DK IVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR  
 orf75a VTAQLLSAYGIQKGLVSVREHNERQMA DK IVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR  
 60 70 80 90 100 110 120

30  
 orf75.pep RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIV  
 orf75a RVREVGFKVVPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLF AKWVRVAFPPV  
 120 130 140 150 160 170 180

35  
 orf75.pep MFETPHRIG AALADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALSADGDQSRGEM  
 orf75a MFETPHRIGATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM  
 180 190 200 210 220 230 240

40  
 orf75.pep VLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALYD  
 orf75a VLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALYDLALSWKNK  
 240 250 260 270 280 290

45  
 orf75a X



The complete length ORF75a nucleotide sequence (SEQ ID NO: 289) is:

```

1  ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
5  151 CCGGTTACCG CGCAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT
201 CAGCGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTTCGGCT
251 ATCTTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGTCCG
351 GTTTAAAGTT GTCCCTGTTG TCGGCGCAAG CGCGGTGATG GCGGCTTTGA
10 401 GTGTGGCTGG TGTGGCGGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TCGGGGTGGC
501 GTTTCCCGTC GTGATGTTTG AAACGCCGCA CCGCATCGGG GCGACGCTTG
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
601 ATCACGAAAA CGTTTGAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
15 651 GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAACATCA TGAAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801 GGAGCTTGCC GCCAAAATCA CGGGCGAGGG AAAAAAAGCT TTGTACGATC
20 851 TGGCACTGTC TTGGA AAAAC AAATGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 290):

```

1  MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQKLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP
25 101 AVCDPGAKLA RRVREVGFV VPVVGASAVM AALSVAGVAG SDFYFNGFVP
151 PKSGERRKLF AKWVRVAFPV VMFETPHRIG ATLADMAELF PERRLMLARE
201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAELA AKITGEGKKA LYDLALS WKN K*

```

ORF75a (SEQ ID NO: 290) and ORF75-1 (SEQ ID NO: 288) show 98.3% identity in 291 aa

30 overlap:

```

10 20 30 40 50 60
orf75a.pep MFQKHLQKASDSVVGGTLYV VATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY
35 orf75-1 MFQKHLQKASDSVVGGTLYV VATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY
10 20 30 40 50 60

70 80 90 100 110 120
orf75a.pep GIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRRVREVGFV
40 orf75-1 GIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRRVREAGFKV
70 80 90 100 110 120

130 140 150 160 170 180
orf75a.pep VPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLF AKWVRVAFPVVMFETPHRIG
45 orf75-1 VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLF AKWVRVAFPIVMFETPHRIG
130 140 150 160 170 180

190 200 210 220 230 240
orf75a.pep m ATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTAL AADGNQSRGEMVLVLYPAQD
50 orf75-1 ATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTAL SADGNQSRGEMVLVLYPAQD
190 200 210 220 230 240

```

		250	260	270	280	290
orf75a.pep	EKHEGLSESAQNIMKILTAELPTKQAAELA	AKITGEGKKALYDLALSWKNKX				
orf75-1	EKHEGLSESAQNIMKILTAELPTKQAAELA	AKITGEGKKALYDLALSWKNKX				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF75 (SEQ ID NO: 286) shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) (SEQ ID NO: 292) from *N. gonorrhoeae*:

10	orf75.pep	MFVFQTAFXMFQKHLQKASDSVVG	GLYVVATPIGNLADITLRALAVLQKA----	AEDTR	56
	orf75ng	MSVFQTAF	FMFQKHLQKASDSVVG	GLYVVATPIGNLADITLRALAVLQKADIICAEDTR	60
	orf75.pep	VTAQLLSAYGIQGLVSVREHNERQ	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR		116
	orf75ng	VTAQLLSAYGIQGLVSVREHNERQ	MADKIVGYLSDGLVVAQVSDAGTPAVCDPGAKLAR		120
15	orf75.pep	RVREAGFKVVPVVGASAVMAALS	VAGVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPV		176
	orf75ng	RVREAGFKVVPVVGASAVMAALS	VAGVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPV		180
	orf75.pep	MFETPHRIGAAALADMAELFPERR	LMLAREITKTFETFLSGTVGEIQTALSADGDQSRGEM		236
20	orf75ng	MFETPHRIGATLADMAELFPERR	LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM		240
	orf75.pep	VLVLYPAQDEKHEGLSESAQNIM	KILTAELPTKQAAELA	AKITGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESAQNAM	KILAAELPTKQAAELA	AKITGEGKKALYDLALSWKNK	300

25 An ORF75ng nucleotide sequence (SEQ ID NO: 291) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 292):

	1	MSVFQTAF	FMFQKHLQKASD	SVVG	GLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGLVSVRE	HNERQMADKV	IGFLSDGLVV	
30	101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES	
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP	
	201	ERRLMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE	
	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK	
	301	*					

35 After further analysis, the following gonococcal DNA sequence (SEQ ID NO: 293) was identified:

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTGCCACGC	CCATCGGCAA	TTTGCCAGAC	ATTACCTGCG
40	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTACAG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
	251	TCCTTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG

451 CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TGCGGGCGGC  
 501 ATTTCTGTGC GTCATGTTTG AAACGCCGCA CCGAATCGGG GCAACGCTTG  
 551 CCGATATGGC GGAATTGTTC CCCGAACGCC GTCTGATGCT GGCGCGCGAA  
 601 ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA  
 651 GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGCGCAG ATGGTGTGG  
 701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCTGCG  
 751 CAAAATGCGA TGAAAATCCT TGCGGCCGAG CTGCCGACCA AGCAGGCGGC  
 801 GGAGCTTGCC GCCAAGATTA CAGGTGAGGG CAAAAGGCT TTGTACGATT  
 851 TGGCACTGTC GTGGAAAAAC AAATGA

This corresponds to the amino acid sequence (SEQ ID NO: 294; ORF75ng-1):

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRLMLARE  
 201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

ORF75ng-1 (SEQ ID NO: 294) and ORF75-1 (SEQ ID NO: 288) show 96.2% identity in 291 aa overlap:

		10	20	30	40	50	60
orf75-1.pep	MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDT	RVTAQLLSAY	
orf75ng-1	MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDT	RVTAQLLSAY	
		10	20	30	40	50	60
orf75-1.pep	GIQKLVSVR	EHNERQMA DK	VIGFLSDGLV	VAQVSDAGTP	AVCDPGAKLA	RRVREAGFKV	
orf75ng-1	GIQKLVSVR	EHNERQMA DK	VIGFLSDGLV	VAQVSDAGTP	AVCDPGAKLA	RRVREAGFKV	
		70	80	90	100	110	120
orf75-1.pep	VPVVGASAVM	AALSVAGVAE	SDFYFNGFVP	PKSGERRKLF	AKWVRAAFPV	VMFETPHRIG	
orf75ng-1	VPVVGASAVM	AALSVAGVAE	SDFYFNGFVP	PKSGERRKLF	AKWVRAAFPV	VMFETPHRIG	
		130	140	150	160	170	180
orf75-1.pep	ATLADMAELF	PERRLMLARE	ITKTFETFLS	GTVGEIQTAL	SADGNQSRGE	MVLVLYPAQD	
orf75ng-1	ATLADMAELF	PERRLMLARE	ITKTFETFLS	GTVGEIQTAL	AADGNQSRGE	MVLVLYPAQD	
		190	200	210	220	230	240
orf75-1.pep	EKHEGLSESA	QNMILKIL	TAELPTKQAA	ELAAKITGEG	KKALYDLAL	SWKNKX	
orf75ng-1	EKHEGLSESA	QNMILKIL	TAELPTKQAA	ELAAKITGEG	KKALYDLAL	SWKNKX	
		250	260	270	280	290	

Furthermore, ORG75ng-1 (SEQ ID NO: 294) shows significant homology to a hypothetical *E.coli* protein (SEQ ID NO: 1131):

```

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
(F286)
)gi|606086 (U18997) ORF_f286 [Escherichia coli]
)gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
5 [Escherichia coli] Length = 286
  Score = 218 bits (550), Expect = 3e-56
  Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

Query: 4   KHLQKASDSVVGGLTYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
10      K  Q  A +S   G LY+V TPIGNLADIT RAL VLQ  D+I AEDTR T  LL  +GI
      Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFGIN 59

Query: 64   GRLVSVREHNERQMADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
      ,      RL ++ +HNE+Q A+ ++  L +G  +A VSDAGTP + DPG  L R  REAG +VVP+
      Sbjct: 60  ARLFALHDHNEQQKAETLLAKLQEQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

Query: 124  VGASAVMAALS VAGVAESDFYFNGFVPPKSGERRKLFKQWVRAAFPVVMFETPHRIGATL 183
15      G  A + ALS AG+   F + GF+P KS  RR      ++ +E+ HR+  +L
      Sbjct: 120 PGPCAAITALSAAGLPDRFCYEGFLPAKSKGRRDALKAEAEPRTLIFYESTHRLDLSL 179

Query: 184  ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
      D+   +  E R ++LARE+TKT+ET      VGE+   +  D N+ +GEMVL++   +
      Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

Query: 243  HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286
20      E  L   A  + +L AELP K+AA LAA+I G  K ALY  AL
      Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 35

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 295):

```

1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
30 51  TTTTGCGGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG .....
      //
651 ..... ..GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 296; ORF76):

```

35 1  MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
      //
201 ..... ..ELVRNQLEQG LRQEKARLKI DALLEENGVK
251 P*

```

40 Further work revealed the complete nucleotide sequence (SEQ ID NO: 297):

```

1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51  TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGG

```

-255-

101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA  
 151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC  
 201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA  
 251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG  
 301 GAGTACGTCC GTTTTCTGGA ACGTTCGAA ACGGTTCCG AAGACGAGCT  
 351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATTG CAGCAGGTCA  
 401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA  
 451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC  
 501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC  
 551 AGTTTGCCGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG  
 601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA  
 651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAGCAG GGTTCGAGAC  
 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC  
 751 AAACCGTAA

This corresponds to the amino acid sequence (SEQ ID NO: 298; ORF76-1):

1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ  
 51 KPDGQAIRND AVRRLQTEV LKNRALKEGL DKDKDVQNRK KIAEASFYAE  
 101 EYVRFLESE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK  
 151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL  
 201 GERYYLFLKS EVGKNPDAQF FELVRNQLEQ GLRQEKARLK IDALLEENGV  
 251 KP\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 (SEQ ID NO: 296) shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) (SEQ ID NO: 300) from strain A of *N. meningitidis*:

		10	20	30		
orf76.pep		MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL				
orf76a		MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND				
		10	20	30	40 50 60	
		//				
			70	80	90	
orf76.pep			XELVRNQLEQGLRQEKARLKIDALLEENGVPKX			
orf76a		DVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLEQGLRQEKARLKIDAILEENGVPKX				
		200	210	220	230 240 250	

The complete length ORF76a nucleotide sequence (SEQ ID NO: 299) is:

1 ATGAAACAGA AAAAAACCGC TGCCGAGTT ATTGCTGCAA TGTTGGCAGG  
 51 TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC  
 101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA  
 151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAC  
 201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA  
 251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG  
 301 GAGTACGTCC GTTTTCTGGA ACGTTCGAA ACGGTTCCG AAAGCGCACT  
 351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA  
 401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA

451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC  
 501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCAGACCG CTGGCTTCGC  
 551 AGTTTGACGC GATGAATCGG GCGGACGTTA CCCGCGATCC GGTCAAATTG  
 601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA  
 651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTTGAGAC  
 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCA TTTTGGAAGA AAACGGTGTC  
 751 AAACCGTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 300):

1 MKQKKTA~~AAV~~ IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ  
 51 KPDGQAIRND AVRRLQTLEV LKNRALKEGL DKDKDVQNR F KIAEASFYAE  
 101 EYVRFLE~~RSE~~ TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK  
 151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL  
 201 GERYYL~~FKLS~~ EVGKNPDAQP FELVRN~~QLEQ~~ GLRQEKARLK IDAILEENG~~V~~  
 251 KP\*

ORF76a (SEQ ID NO: 300) and ORF76-1 (SEQ ID NO: 298) show 97.6% identity in 252 aa overlap:

20	orf76a.pep	10 20 30 40 50 60	MKQKKTA <del>AAV</del> IAAMLAGFAA AKAPEIDPALVDTLVAQIMQ QADRHAEQSQ KPDGQAIRND
	orf76-1	10 20 30 40 50 60	MKQKKTA <del>AAV</del> IAAMLAGFAA AKAPEIDPALVDTLVAQIMQ QADRHAEQSQ KPDGQAIRND
25	orf76a.pep	70 80 90 100 110 120	AVRRLQTLEV LKNRALKEGL DKDKDVQNR FKIAEASFYAE EYVRFLE RSETVSESALRQF
	orf76-1	70 80 90 100 110 120	AVRRLQTLEV LKNRALKEGL DKDKDVQNR FKIAEASFYAE EYVRFLE RSETVSEDELHKF
30	orf76a.pep	130 140 150 160 170 180	YERQIRMIKL QQVSFATEEE ARQAQQLLLK GLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
	orf76-1	130 140 150 160 170 180	YEQQIRMIKL QQVSFATEEE ARQAQQLLLK GLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
35	orf76a.pep	190 200 210 220 230 240	LASQFAAMNR GDVTRDPVKLGERYYL FKLSEVGKNPDAQPFELVRN <del>QLEQ</del> GLRQEKARLK
	orf76-1	190 200 210 220 230 240	LASQFAAMNR GDVTRDPVKLGERYYL FKLSEVGKNPDAQPFELVRN <del>QLEQ</del> GLRQEKARLK
40	orf76a.pep	250	IDAILEENG <del>V</del> KPX
	orf76-1	250	IDALLEENG <del>V</del> KPX

Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 (SEQ ID NO: 296) and a predicted ORF (ORF76.ng) (SEQ ID NO: 302) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

```

5      orf76.pep      MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL      30
      |||||||||||||||||||
      orf76ng        MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND      60
                        //
      orf76.pep                        ELVRNQLEQGLRQEKARLKIDALLEENGVKP      251
      |||||||||||||||||||
10     orf76ng        VTRNPVKLGERYYLFKLGAVGKNPDAQPFELVRNQLEQGLRQEKARLKIDALLEENGVKP      251

```

The complete length ORF76ng nucleotide sequence (SEQ ID NO: 301) is:

```

1      ATGAAACAGA AAAAGACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
51     TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
15     101    TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
      151    AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAC
      201    TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
      251    AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
20     301    GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
      351    GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
      401    GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
      451    GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
25     501    GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTcgc
      551    agtttgCCGG TATGAACCGT GGCGACGTTA CCCGCAATCC GGTCAAATTG
      601    GGCGAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
      651    CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTTGAGGC
      701    AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA Aacgggtgtc
      751    AaacCGTAA

```

30 This encodes a protein having amino acid sequence (SEQ ID NO: 302):

```

1      MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51     RPDGQAIRND AVRRLQTLEV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
35     101    EYVRFLESE TVSESALRQF YERQIRMIKL QVVSFATEEE ARQAQQLLLK
      151    GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
      201    GERYYLFLKG AVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENGV
      251    KP*

```

ORF76ng (SEQ ID NO: 302) and ORF76-1 (SEQ ID NO: 298) show 96.0% identity in 252 aa overlap

```

40     10      20      30      40      50      60
      orf76-1.pep  MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      |||||||||||||||||||
      orf76ng      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQRPDGQAIRND
                        10      20      30      40      50      60

45     70      80      90      100     110     120
      orf76-1.pep  AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSEDELHKF
      |||||||||||||||||||
      orf76ng      AVRRLQTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSESALRQF

```

-258-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf76-1.pep	YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP					
	orf76ng	YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP					
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf76-1.pep	LASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLEQGLRQEKARLK					
	orf76ng	LASQFAGMNRGDVTRDPVKLGERYYLFLKGA VGKNPDAQPFELVRNQLEQGLRQEKARLK					
		190	200	210	220	230	240
		250					
15	orf76-1.pep	IDALLEENGVKPX					
	orf76ng	IDALLEENGVKPX					
		250					

Furthermore, ORF76ng (SEQ ID NO: 302) shows significant homology to a *B.subtilis* export protein precursor (SEQ ID NO: 1132):

20	sp P24327 PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR )gi 98227 pir  S15269	
	33K lipoprotein - Bacillus subtilis )gi 39782 (X57271) 33kDa lipoprotein [Bacillus subtilis]	
	)gi 2226124 gnl PID e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]	
25	)gi 2633331 gnl PID e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]	
	Length = 292	
	Score = 50.4 bits (118), Expect = 1e-05	
	Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)	
30	Query: 70	VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAE EYVRFLERSETVSE 114
		VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
	Sbjct: 53	VLTQLVQEKVLDKKYKVS DKEIDNKLKEYKTQLGDQYTALEKQY GKDYLKEQVKYELLTQ 112
	Query: 115	SA-----LRQFYERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPN 163
		A +++++E I+ + A ++ A + ++ L KG FE L K Y
	Sbjct: 113	KAADNIKVT DADIKEYW E LK GKIRASHILVADKKTAE E V E K L K K G E K F E D L A K E Y S T 172
35	Query: 164	DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDA 218
		D A G F Q+ E + + G+V+ DPVK Y++ K +E D
	Sbjct: 173	DSSASKGGDLGWFAKEGQMD E T F S K A A F L K T G E V S - D P V K T Q Y G Y H I I K K T E E R G K Y D D 231
	Query: 219	QPFELVRNQLEQGLRQEKA 237
		EL LEQ L A
	Sbjct: 232	MKKELKSEVLEQKLNDNAA 250

- 40 Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.



ORF76-1 (SEQ ID NO: 298) (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 (SEQ ID NO: 298) is a surface-exposed protein, and that it is a useful immunogen.

### Example 36

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 303):

```

10      1  ATGAAAAAAT CTTTCCTTAC GCTTGTCTCG TATTCGTCTT TACTTACCGC
      51  CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
     101  GCAAAAATTG CGGAAACGTT TGCGCTGACA TTTGTGATTG CTGCGCTGTA
     151  TCTGTTTGGC CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
     201  CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
     251  ATGACG.... //
     15      1201 ..... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
     1251  ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
     1301  ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
     1351  GTTGTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
     20      1401  CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTCAACGTT CCTGATTCAC
     1451  ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTCGCGAAG GCTCGGTAAC
     1501  GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
     1551  AGGCGGAATA TGTTTATCCG CAATGA

```

25 This corresponds to the amino acid sequence (SEQ ID NO: 304; ORF81):

```

30      1  MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAKIAETF ALTFVIAALY
     51  LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT..... //
     401  ...QTVFEQL QKTPDGNWLF AYTSDHGQYV RQDIYNQGT V QPDSYLVPLV
     451  LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
     501  GNLTGDAGS LNIRDGKAEY VYPQ*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 305):

```

35      1  ATGAAAAAAT CTTTCCTTAC GCTTGTCTCG TATTCGTCTT TACTTACCGC
     51  CAGCGAAATT GCCTATCGCT TTGTATTTGG GATTGAAACC TTACCGGCGG
     101  CAAAATTGCG GGAACGTTT GCGCTGACAT TTGTGATTGC TGCGCTGTAT
     151  CTGTTTTCGC GTTATAAGGT GACGCGTTTG TTGATTGCGG TGTTTTTTGC
     201  GTTCAGCATT ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
     251  TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
     40      301  AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGGCTGCCTG TGTGTGGGG
     351  CGTGTTGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
     401  CGCATTTTTC TGCCGATATA CTGTTTGCC TCCTAATGCT GATGATTTTC
     451  GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
     501  ATACAGCCGC ATCAAAGCCA ATTATTTTCAG CTTCGGTTAT TTTGTCGGAC
     45      551  GCGTGTTGCC GTATCAGTTG TTTGATTAA GCAGGATTCC CGCCTTTAAG

```

5  
10  
15  
20

```

601 CAGCCTGCTC CAAGCAAAAT CGGGCAGGGC AGTGTTCAAA ATATCGTCCT
651 GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAGCTG TTTGGCTACG
701 GACGCGAAAC TTCGCCGTTT TTAACCCGGC TGTCGCAAGC CGATTTTAAG
751 CCGATTGTGA AACAAAGTTA TTCCGCAGGC TTTATGACTG CAGTGTCCCT
801 GCCCAGTTTT TTCAATGCGA TACCGCACGC CAACGGCTTG GAACAAATCA
851 GCGGCGGCGA TACCAATATG TTCCGCCTCG CCAAAGAGCA GGGCTATGAA
901 ACGTATTTTT ACAGCGCGCA GCGGAAAAAC GAGATGGCGA TTTTGAACCT
951 AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
1001 ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
1051 AAAATCAATT TGCAGCAGGG CAAGCATTTT ATCGTGTTCG ACCAACGCGG
1101 TTCGCACGCC CCATACGGCG CATTGTTGCA GCCTCAAGAT AAAGTATTCTG
1151 GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTT CGCCAAGATA
1301 TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATCTCGT GCCGCTAGTG
1351 TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
1401 GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTC CTGATTACACA
1451 CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACG
1501 GGCAACCTGA TTACGGGTGA TGCAGGCAGC TTGAACATTC GCGACGGCAA
1551 GGCGGAATAT GTTTATCCGC AATGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 306; ORF81-1):

25  
30

```

1 MKKSFLTLVL YSLLTASEI AYRFVFGIET LPAAKIAETF ALTFVIAALY
51 LFARYKVTRL LIAVFFAFSI IANNVHYAVY QSWMTGINYW LMLKEVTEVG
101 SAGASMLDKL WLPVLWGVLE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSPAPAFK
201 QPAPSKIGQG SVQNIIVLIM ESESAHLKL FGYGRETSPF LTRLSQADFK
251 PIVKQSYSAG FMTAVSLPSF FNAIPHANGL EQISGGDTNM FRLAKEQGYE
301 TYFYSAQAEEN EMAILNLIGK KWIDHLIQPT QLGYNNGDNM PDEKLLPLFD
351 KINLQQKHF IVLHQRGSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
401 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTQ QPDSYLVPLV
451 LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGVDMP VSGCREGSVT
501 GNLITGDAGS LNIRDGKAEY VYPQ*

```

35 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 (SEQ ID NO: 304) shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) (SEQ ID NO: 308) from strain A of *N. meningitidis*:

40  
45  
50

```

              10      20      30      40      50      60
orf81.pep    MKKSFLTLVL YSLLTASEI AYPLELGIET LPAAKIAETF ALTFVIAALY LFARNKVTRL
              |||:::| ||||| ||||| : : ||||| ||||| ||||| ||||| |||||
orf81a       MKKSLFVLFL YSLLTASEI AYRFVFGIET LPAAKMAETF ALTFVIAALY LFARYKATRL
              10      20      30      40      50      60

              70      80
orf81.pep    LIAVFFAFSI IANNVHYADY QSWMT
              ||||| ||||| ||||| ||||| :
orf81a       LIAVFFAFSI IANNVHYAVY QSWITGINYW LMLKEITEVGGAGASMLDKLWLPALWGVLE
              70      80      90      100     110     120
              //
                      120     130     140

```

-261-

	orf81.pep																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
--	-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

The complete length ORF81a nucleotide sequence (SEQ ID NO: 307) is:

	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTCTG	TATTCGTCCC	TACTTACTGC
	51	CAGCGAAATT	GCTTATCGCT	TTGTATTCGG	AATTGAAACC	TTACCGGCTG
20	101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGGC	AACGCGTTTG	TTGATTGCGG	TGTTTTTCGC
	201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
	251	TAACGGGCAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
25	351	CGTGTGGAA	GTCATGTTGT	TTTGCAGCCT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTC	TGCCGATATA	CTGTTGCCT	TCCTAATGCT	GATGATTTTC
	451	GTGCGTTCGT	TCGACACGAA	ACAAGAACAC	GGTATTTTCG	CCAAACCGAC
	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTTCT	CTTCGGTTAT	TTTGTCCGAC
	551	GCGTGTGCGC	GTATCAGTTG	TTTGATTAA	GCAAGATTCC	TGTGTTCAAA
	601	CAGCTTGCTC	CAAGCAGAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCCT
30	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGCTACG
	701	GGCGCGAAAC	TTCGCCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAAAG
	751	CCGATTGTGA	AACAAAGTTA	TTCCGCAGGC	TTTATGACGG	CAGTATCCCT
	801	GCCCAGTTTC	TTTAACGTCA	TACCGCATGC	CAACGGCTTG	GAACAAATCA
35	851	GCGGCGGCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
	901	CAAAATGATT	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
	951	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTT	CGCCAAGATA
	1001	TCTACAATCA	AGGCACGGTG	CAGCCCGACA	GCTATCTCGT	GCCGCTGGTG
	1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCCAACC	AGGCTTTTGC
40	1101	GCCTTGCGAG	ATTGCCCTTC	ATCAGCAGCT	TTCAACGTTT	CTGATTACACA
	1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAACG
	1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTC	GCGACGGCAA
	1251	GGCGGAATAT	GTTTATCCGC	AATGA		

This encodes a protein having amino acid sequence (SEQ ID NO: 308):

45	1	MKKSFLVFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFVIAALY
	51	LFARYKATRL	LIAVFFAFSI	IANNVHYAVY	QSWITGINYW	LMLEITEVVG
	101	GAGASMLDKL	WLPALWGVLE	VMLFCSLAKF	RRKTHFSADI	LFAFLMLMIF
	151	YRSFDTKQEH	GISPKPTYSR	IKANYFSFGY	FVGRVLPYQL	FDLSKIPVFK
	201	QPAPSRIGQG	SIQNIIVLIM	ESESAAHLKL	FGYGRETSPF	LTQLSQADFK
50	251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANGL	EQISGGDIVD	KYDNTIHKTD
	301	QMIQTVFEQL	QKQPDGNWLF	AYTSDHGQYV	RQDIYNQGT	VPDSYLVPLV
	351	LYSPDKAVQQ	AANQAFAPCE	IAFHQQLSTF	LIHTLGVDMP	VSGCREGSVT
	401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

ORF81a (SEQ ID NO: 308) and ORF81-1 (SEQ ID NO: 306) show 77.9% identity in 524 aa overlap:

```

5      orf81a.pep  10      20      30      40      50      60
      orf81-1      MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFVIAALYLFARYKATRL
      MKKSFLTTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
      10      20      30      40      50      60

10     orf81a.pep  70      80      90      100     110     120
      orf81-1      LIAVFFAFSIIANNVHYAVYQSWITGINYWMLMLKEITEVGGAGASMLDKLWLPALWGVLE
      LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLMLKEVTEVGSAGASMLDKLWLPVLWGVLE
      70      80      90      100     110     120

15     orf81a.pep  130     140     150     160     170     180
      orf81-1      VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
      VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
      130     140     150     160     170     180

20     orf81a.pep  190     200     210     220     230     240
      orf81-1      FVGRVLPYQLFDLSKIPVFKQPAPSRIGQGSIQNIVLIMGESESA AHLKLFYGRGRETSPF
      FVGRVLPYQLFDLSRIPAFKQPAPSKIQQGSVQNIIVLIMGESESA AHLKLFYGRGRETSPF
      190     200     210     220     230     240

25     orf81a.pep  250     260     270     280
      orf81-1      LTQLSQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGD-----
      LTRLNQADFKPIVKQSYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE
      250     260     270     280     290     300

30     orf81a.pep  -----
      orf81-1      TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF
      310     320     330     340     350     360

35     orf81a.pep  ----- 290     300     310     320
      orf81-1      IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF
      370     380     390     400     410     420

40     orf81a.pep  330     340     350     360     370     380
      orf81-1      AYTSDHGQYVRQDIYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF
      AYTSDHGQYVRQDIYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF
      430     440     450     460     470     480

45     orf81a.pep  390     400     410     420
      orf81-1      LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
      LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
      490     500     510     520

```

Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 (SEQ ID NO: 304) and a predicted ORF (ORF81.ng) (SEQ ID NO: 310) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

5	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTFFVIAALYLFARNKVTRL	60
	orf81ng	MKKSFLVLFYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFFMIAALYLFARYKASRL	60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
10	orf81ng	LIAVFFAFSMIANNVHYAVYQSWMTGINYWMLLKEVTEVGSAGASMLDKLWLPALWGVAE	120
	orf81.pep	QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD	433
15	orf81.pep	IYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81ng	IYNQGTVPDPSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81.pep	CREGSVTGNLITGDAGSLNIRDGKAELYVYPQ	524
20	orf81ng	CREGSVTGNLITGDAGSLNIRNGKAELYVYPQ	524

The complete length ORF81ng nucleotide sequence (SEQ ID NO: 309) is:

	1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCATCCC	TACTTACCGC
25	51	CAGCGAAATC	GCCTATCGCT	TTGTATTCGG	AATTGAAACC	TTACCGGCTG
	101	CAAAAATGGC	GGAAACGTTT	GCGCTGACAT	TTATGATTGC	TGCGCTGTAT
	151	CTGTTTGC	GTTATAAGGC	TTCGCGGCTG	CTGATTGCGG	TGTTTTTCGC
	201	GTTCAAGCATG	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
	251	TGACGGGTAT	TAATATTGG	CTGATGCTGA	AAGAGTTAC	CGAAGTCGGC
30	301	AGCGCGGGCG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CTTTGTGGGG
	351	CGTGCGGAA	GTCATGTTGT	TTTGACGCTT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTC	TGCCGATATA	CTGTTTGCTT	TCCTAATGCT	GATGATTTTC
	451	GTGCGTTCGT	TCGACACGAA	ACAAGAGCAC	GGTATTTTCG	CCAAACCGAC
	501	ATACAGCCGC	ATCAAAGCCA	ATTATTTTCAG	CTTCGGTTAT	TTTGTCGGGC
	551	GCGTGTTGCC	GTATCAGTTG	TTTGATTAA	GCAAGATCCC	TGTGTTCAAA
35	601	CAGCCTGCTC	CAAGCAAAAT	CGGGCAAGGC	AGTATTCAA	ATATCGTCTC
	651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGTTACG
	701	GGCGCGAAAC	TTCGCCGTTT	TTAACCCGGC	TGTCGCAAGC	CGATTTTAAG
	751	CCGATTGTGA	AACAAAGTTA	TTCCGCAGGC	TTTATGACGG	CAGTATCCCT
	801	GCCCAGTTTC	TTTAACGTCA	TACCGCACGC	CAACGGCTTG	GAACAAATCA
40	851	GCGGCGGCGA	TACCAATATG	TTCCGCCTCG	CCAAAGAGCA	GGGCTATGAA
	901	ACGTATTTTT	ACAGTGCCCA	GGCTGAAAAC	CAAATGGCAA	TTTTGAACTT
	951	AATCGGTAAG	AAATGGATAG	ACCATCTGAT	TCAGCCGACG	CAACTTGGCT
	1001	ACGGCAACGG	CGACAATATG	CCCGATGAGA	AGCTGCTGCC	GTTGTTTCGAC
45	1051	AAAATCAATT	TGCAGCAGGG	CAGGCATTTT	ATCGTGTTGC	ACCAACGCGG
	1101	TTCCGACGCC	CCATACGGCG	CATTGTTGCA	GCCTCAAGAT	AAAGTATTCG
	1151	GCGAAGCCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
	1201	CAAATGATTC	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
	1251	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTG	CGCCAAGATA
	1301	TCTACAATCA	AGGCACGGTG	CAGCCCGACA	GCTATATTGT	GCCTCTGGTT

1351 TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC  
 1401 GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTTCACA  
 1451 CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACA  
 1501 GGCAACCTGA TTACGGGCGA TGCAGGCAGC TTGAACATTC GCAACGGCAA  
 1551 GGCGGAATAT GTTTATCCGC AATAA

This encodes a protein having amino acid sequence (SEQ ID NO: 310):

1 MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFMIAALY  
 51 LFARYKASRL LIAVFFAFSM IANNVHYAVY QSWMTGINYW LMLKEVTEVG  
 101 SAGASMLDKL WLPALWGVAE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF  
 151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLSKIPVFK  
 201 QPAPSKIGQG SIQNIIVLIMG ESESAHLKL FGYGRETSPF LTRLSQADFK  
 251 PIVKQSYSAG FMTAVSLPSF FNVIPHANGL EQISGGDTNM FRLAKEQGYE  
 301 TYFYSAQAEN QMAILNLIGK KWIDHLIQPT QLGYGNGDNM PDEKLLPLFD  
 351 KINLQQGRHF IVLHQRGSHA PYGALLQPQD KVFGEADIVD KYDNTIHKTD  
 401 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYIVPLV  
 451 LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT  
 501 GNLTG DAGS LNIRNGKAEY VYPQ\*

ORF81ng (SEQ ID NO: 310) and ORF81-1 (SEQ ID NO: 306) show 96.4% identity in 524 aa overlap:

		10	20	30	40	50	60
orf81ng-1.pep		MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL					
		::			:	:	:
orf81-1		MKKSFLLTVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf81ng-1.pep		LIAVFFAFSMIANNVHYAVYQSWMTGINYWMLMLKEVTEVGSAGASMLDKLWLPALWGVAE					
		:				:	
orf81-1		LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLMLKEVTEVGSAGASMLDKLWLPVLWGVAE					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf81ng-1.pep		VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR IKANYFSFGY					
orf81-1		VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR IKANYFSFGY					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf81ng-1.pep		FVGRVLPYQLFDLSKIPVFKQPAPSKIGQGSIQNIIVLIMGESESAHLKLFGYGRETSPF					
			:	:	:	:	:
orf81-1		FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFGYGRETSPF					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf81ng-1.pep		LTRLSQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGDTNMFRLAKEQGYE					
			:	:	:	:	:
orf81-1		LTRLSQADFKPIVKQSYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE					
		250	260	270	280	290	300
		310	320	330	340	350	360
orf81ng-1.pep		TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGRHF					
							:
orf81-1		TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					

-265-

		310	320	330	340	350	360
		370	380	390	400	410	420
	orf81ng-1.pep	IVLHQRGSHAPYGALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF					
5	orf81-1	IVLHQRGSHAPYGALLQPQDKVFGEADIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLF					
		370	380	390	400	410	420
		430	440	450	460	470	480
	orf81ng-1.pep	AYTSDHGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
10	orf81-1	AYTSDHGQYVRQDIYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
		490	500	510	520		
	orf81ng-1.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRNGKAEYVYPQX					
15	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		

Furthermore, ORF81ng (SEQ ID NO: 310) shows significant homology to an *E.coli* OMP (SEQ ID NO: 1133):

20	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)	
25	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMQRLFFVLTLVVKRISSLPLRLLVAAPFVL-LTAADMSISLY- 86	
	Query: 82 SWMT-----GINYWLMLKEVTEVGSAGASMLDKLWLPALWGVAEVMLFCSLAKFRRT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLAVIIKYDV 141	
30	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLLIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201	
35	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPYFQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257	
	Query: 242 TRLSQADFKPIVKQSYSAGFMTAVSLP---SFFNVIPHANGLEQISGGDTNMFRLAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPLSLTADSVLSH-----DIHNPDPNIINMANQAG 310	
40	Query: 299 YETYFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLLPHLSQLALQQ 359	
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPQDKVFGEADIVDK-YDNTIHKTDQMIQTVFEQLQK 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPQDDQDACYDNSIHYTDSLGLGQVFELLK- 418	
45	Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPDSYIVPL-VLYSP 454	

D            Y +DHG            ++++Y G            +Y VP+ + YSP  
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 37

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 311):

```

10      1  ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGCG GCACACTGAC
      51  CGGCATACTC GCCCaCGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
     101  TCGTCGCCCGC ATCGTCCC GC CGCCGTCA AAGAAATGGA TTTGTCCGCC
     151  yTAAAAGGAC GCAAAGCCGC CyTTTACGTC TCCGTTATGG GCGACCAAGG
     201  TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
     251  GCGGCTACCA CAACAACCCC GAAAGTGCCA CCCAATACAG CTACCCCGCC
     15  301  TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
     351  TTCCACATCG CTTTGAACG CCCCCGCCGC CGyCyTGACG AAAAACAGCG
     401  GACGCAAAGG CGAACGcTCC GCCGACTGT CCGTCAACGG CACGGGCGAC
     451  TACGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
     501  CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
     20  551  CCGrATACGC CGACACCGAC GTATTCTGTA CCGTCGACGT A...
```

This corresponds to the amino acid sequence (SEQ ID NO: 312; ORF83):

```

25      1  ..TLLFIPLVL TXCGTLTGIL AHGGGKRFV EQELVAASSR AAVKEMDLA
     51  LKGRKAAXYV SVMGDQSGN ISGGYRYSIDA LIRGGYHNNP ESATQYSYPA
     101  YDTTATTKSD ALSSVTTSTS LLNAPAAXLT KNSGRKGRS AGLSVNGTGD
     151  YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVTVDV..
```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 313):

```

30      1  ATGAAAACCC TGCTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
     51  ACTGACCGGC ATACCCGCCC ACGGCGGCGG CAAACGCTTT GCCGTCGAAC
     101  AAGAACTCGT CGCCGCATCG TCCCGCGCCG CCGTCAAAGA AATGGATTTG
     151  TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
     201  CCAAGTTTCG GGCAACATAA GCGGCGGACG CTACTCTATC GACGCACTGA
     251  TACGCGGCGG CTACCACAAC AACCCCGAAA GTGCCACCCA ATACAGCTAC
     35  301  CCCGCCTACG AACTACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
     351  AACCCTTCC ACATCGCTTT TGAACGCCCC CGCCGCCGCC CTGACGAAAA
     401  ACAGCGGACG CAAAGGCGAA CGCTCCGCCG GACTGTCCGT CAACGGCACG
     451  GGCGACTACC GCAACGAAAC CCTGCTCGCC AACCCCGCG ACGTTTCCTT
     501  CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
     40  551  TACCGCCCGA ATACGCGGAC ACCGACGTAT TCGTAACCGT CGACGTATTG
     601  GGCACCGTCC GCAGCCGTAC CGAACTGCAC CTCTACAACG CCGAAACCTT
     651  TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCCGGA
     701  AACTGTGAT TACCCCTAAA ACCGCCGCTT ACGAATCCCA ATACCAAGAA
     751  CAATACGCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
     45  801  CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATTACCCCC TACGGCGACA
```



851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCC  
 901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This corresponds to the amino acid sequence (SEQ ID NO: 314; ORF83-1):

5           1 MKTLLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL  
           51 SALKGRKAAL YVSMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY  
          101 PAYDTTATTK SDALSSVTTS TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT  
          151 GDYRNETLLA NPRDVSFLT NLIQTVFYLRG IEVVPPEYAD TDVFVTVDVF  
 10          201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE  
          251 QYALWTGPYK VSKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNKKPK  
          301 DVGNEVIRRR KGG\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF83 (SEQ ID NO: 312) shows 96.4% identity over a 197aa overlap with an ORF (ORF83a)  
 (SEQ ID NO: 316) from strain A of *N. meningitidis*:

		10	20	30	40	50
orf83.pep		TLLLFIPVLVTXCGTLTGILAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAX				
		:				
20	orf83a	MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL				
		10	20	30	40	50
		60	70	80	90	100
orf83.pep		YVSMGDQGS GNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS				
25	orf83a	YVSMGDQGS GNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS				
		70	80	90	100	110
		120	130	140	150	160
orf83.pep		TSLLNAPAAAXLTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLT NLIQTVFYLRG				
30	orf83a	TSLLNAPAAALTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLT NLIQTVFYLRG				
		130	140	150	160	170
		180	190			
orf83.pep		IEVVPXYADTDVFVTVDV				
35	orf83a	IEVVPPEYADTDVFVTVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK				
		190	200	210	220	230
						240

The complete length ORF83a nucleotide sequence (SEQ ID NO: 315) is:

40           1 ATGAAAACCC TGCTCNTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC  
           51 ACTGACCGGC ATACCGGCC ACGGCGGCGG CAAACGCTTT GCCGTGGAAC  
          101 AAGAACTCGT CGCCGCATCG TCCCGCGCCG CCGTCAAAGA AATGGACTTG  
          151 TCCGCCCTGA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA  
          201 CCAAGGTTCTG GGCAACATAA GCGGCGGACG CTACTCTATC GACGCACTGA  
          251 TACGCGGCGG CTACCACAAC AACCCGAAA GTGCCACCCA ATACAGCTAC  
 45          301 CCCGCTACG ACATACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT  
          351 AACCACTTCC ACATCGCTTT TGAACGCCCC CGCCGCCGCC CTGACGAAAA

5  
10  
401 ACAGCGGACG CAAAGGCGAA CGCTCCGCCG GACTGTCCGT CAACGGCAGC  
451 GGC GACTACC GCAACGAAAC CCTGCTCGCC AACCCCGCG ACGTTTCCTT  
501 CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG  
551 TACCGCCCGA ATACGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTC  
601 GGCACCGTCC GCAGCCGCAC CGAACTGCAC CTCTACAACG CCGAAACCCCT  
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCCGGA  
701 AACTGCTGAT TGCCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA  
751 CAATACGCCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC  
801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCCC TACGGCGACA  
851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAAACCC  
901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence (SEQ ID NO: 316):

15  
20  
1 MKTLLXLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL  
51 SALKGRKAAL YVSVMGDQGS GNISGGYRSI DALIRGGYHN NPESATQYSY  
101 PAYDTTATTK SDALSSVTTS TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT  
151 GDYRNETLLA NPRDVSFLT NLIQTVFYLRG IEVVPPEYAD TDVFVTVDFV  
201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE  
251 QYALWMPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRD DFKQNNGKKP  
301 DVGNEVIRRR KGG\*

ORF83a (SEQ ID NO: 316) and ORF83-1 (SEQ ID NO: 314) show 98.4% identity in 313 aa overlap:

25  
30  
35  
40  
45  
50  
orf83a.pep MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL  
orf83-1 MKTLLLLLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL  
orf83a.pep YVSVMGDQGS GNISGGYRSI DALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS  
orf83-1 YVSVMGDQGS GNISGGYRSI DALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS  
orf83a.pep TSLLNAPAAALTKN SGRK GERSAGLSVNGTGDYRNETLLANPRDVSFLT NLIQTVFYLRG  
orf83-1 TSLLNAPAAALTKN SGRK GERSAGLSVNGTGDYRNETLLANPRDVSFLT NLIQTVFYLRG  
orf83a.pep IEVVPPEYADTDVFVTVDFVGT VRSRTELHLYNAETLKAQTKLEYFAVDR DSRKLLIAPK  
orf83-1 IEVVPPEYADTDVFVTVDFVGT VRSRTELHLYNAETLKAQTKLEYFAVDR DSRKLLITPK  
orf83a.pep TAAYESQYQEYALWMPYSVGKTVKASDR LMVDFSDITPYGDTTAQNRD DFKQNNGKKP  
orf83-1 TAAYESQYQEYALWMPYSVGKTVKASDR LMVDFSDITPYGDTTAQNRD DFKQNNGKKP  
orf83a.pep DVGNEVIRRRKGGX

orf83-1            |||||  
                   DVGNEVIRRRKGGX  
                   310

Homology with a predicted ORF from *N.gonorrhoeae*

- 5 ORF83 (SEQ ID NO: 312) shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) (SEQ ID NO: 318) from *N. gonorrhoeae*:

	orf83.pep	TLLLFIPVLVTXCGTLTGILAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAX	58
	orf83ng	MKTLLLLLIPLVLTACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL	60
10	orf83.pep	YVSVMGDQGSNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS	118
	orf83ng	YVSVMGDQGSNISGGRYSIDALIRGGYHNNPDSATRYSPAYDTTATTKSDALSGVTTS	120
	orf83.pep	TSLLNAPAAXLTKNSGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG	178
15	orf83ng	TSLLNAPAAALTKNNGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG	180
	orf83.pep	IEVPPXYADTDVFVTVDV	197
	orf83ng	IEVPPPEYADTDVFVTVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK	240

- 20 The complete length ORF83ng nucleotide sequence (SEQ ID NO: 317) is:

	1	ATGAAAACCC	TGCTCCTCCT	CATCCCCCTC	GTACTCACCG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCCC	ACGGCGGCGG	CAAACGCTTT	GCCGTGGAAC
	101	AGGAACTCGT	CGCCGCATCG	TCCCGCGCCG	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGCAA	AGCCGCCCTT	TACGTCTCCG	TTATGGGCGA
25	201	CCAAGGTTTC	GGCAACATAA	GCGGCGGACG	CTACTCCATC	GACGCACTGA
	251	TACGCGGCGG	CTACCACAAC	AACCCCGACA	GCGCCACCCG	ATACAGCTAC
	301	CCCGCCTATG	ACACTACCGC	CACCACCAAA	TCCGACGCGC	TCTCCGGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCCGCC	CTGACGAAAA
	401	ACAACGGACG	CAAAGGCGAA	CGTCCGCCG	GACTGTCCGT	CAACGCGACG
30	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCG	ACGTTTCCTT
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CCTGCGCGGC	ATCGAAGTCG
	551	TACCGCCCGA	ATACGCCGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTC
	601	GGCACCGTCC	GCAGCCGTAC	CGAACTGCAC	CTCTACAACG	CCGAAACCTT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTTCG	CGTCGACCGC	GACAGCCGGA
35	701	AACTGCTGAT	TGCCCCTAAA	ACCGCCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCC	TCTGGATGGG	ACCTTACAGC	GTCGGCAAAA	CCGTCAAAGC
	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCCC	TACGGCGACA
	851	CAACCGCCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACGG	TAAAAACCCC
40	901	GATGTCGGCA	ACGAAGTCAT	CCGCCGCCG	AAAGGAGGAT	AA

This encodes a protein having amino acid sequence (SEQ ID NO: 318):

	1	MKTLLLLLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGRYSI	DALIRGGYHN	NPDSATRYSY
	101	PAYDTTATTK	SDALSGVTTS	TSLLNAPAAA	LTKNNGRKGE	RSAGLSVNGT
45	151	GDYRNETLLA	NPRDVSFLTNI	LIQTVFYLRG	IEVPPPEYAD	TDVFVTVDVF
	201	GTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLLIAPK	TAAYESQYQE
	251	QYALWMPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNNGNPN

301 DVGNEVIRRR KGG\*

ORF83ng (SEQ ID NO: 318) and ORF83-1 (SEQ ID NO: 314) show 97.1% identity in 313 aa overlap

```
5      10      20      30      40      50      60
orf83-1.pep MKTLLLLIPLVLTACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL
orf83ng      MKTLLLLIPLVLTACGTLTGIPAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL
           10      20      30      40      50      60

10     70     80     90     100    110    120
orf83-1.pep YVSMGDQGSNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS
orf83ng      YVSMGDQGSNISGGGRYSIDALIRGGYHNNPDSATRYSPAYDTTATTKSDALSGVTTS
           70     80     90     100    110    120

15     130    140    150    160    170    180
orf83-1.pep TSLLNAPAAALTKNSGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG
orf83ng      TSLLNAPAAALTKNNGRKGERSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG
           130    140    150    160    170    180

20     190    200    210    220    230    240
orf83-1.pep IEVVPPEYADTDVFTVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLITPK
orf83ng      IEVVPPEYADTDVFTVDVFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLLIAPK
           190    200    210    220    230    240

25     250    260    270    280    290    300
orf83-1.pep TAAYESQYQEYALWTGPYKVSTVKASDRLMVDFSDITPYGDTTAQNRPDFKQNNGKKP
orf83ng      TAAYESQYQEYALWMGPYSVGKTVSKASDRLMVDFSDITPYGDTTAQNRPDFKQNNGKKNP
           250    260    270    280    290    300

30     310
orf83-1.pep DVGNEVIRRRKGGX
orf83ng      DVGNEVIRRRKGGX
           310
```

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 319):

```

      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
      5  51  AAAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
      10 101  AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAAATACCG
      15 151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
      20 201  GCAGCTTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
      25 251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
      30 301  TCGGCAGGTT CAAAAATCCC TGAATATGTC CAATGGCTGA ATACGCACAG
      35 351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
      40 401  ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACAT CGCTTCAAAC
      45 451  AAGATGGGTA TCGGTACGCT TTTAGAAATGG AAAATATGCG CGGACGATCC
      50 501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
      55 551  AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
      60 601  AAGCGGTCAA AGTGGTTTTC CACTCTGCCA GTAATAGTAT TGCTGATTCC
      65 651  CGTGTGTTGTC GGCCTGTCCT ATAAAATGTT GagCaGTTAC GGAAAAAAC
      70 701  aGGAAGAACC CGCAGCACA GAATCGGCGG CAACAGAACA GCAGGCAGTA
      75 751  CTTCCCGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
      80 801  AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCcgATTT
      85 851  ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
      90 901  GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
      95 951  gaAAGAAGTG ACGGaGTTGA TGTGcgaAgG aCTATGTaAA AAacGGCTTG
     100 1001 CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGGCAGGAAG TTCAGCAAAG
     105 1051 CGCGCAgCAA CATTCGACAA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
     110 1101 CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACGCG GGAAACCGTT
     115 1151 TGAAGGAATC GGaCGGGGGC GTGGTCCGAT CGGCAAACTG A

```

This corresponds to the amino acid sequence (SEQ ID NO: 320; ORF84):

```

      1  MAEICLITGT PGSGKTLKMV SMMANDEMFK PDEKAIRRKV FTNIKGLKIP
     30  51  HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVPWPAR
     60 101  SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
     90 151  KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNVK
    120 201  KRSKWFTYLP VIVLLIPVFV GLSYKMLSSY GKKQEPEPAAQ ESAATEQQAV
    150 251  LPDKTEGEPV NNGNLTADMF VPTLSEKPKS KPIYNGVRQV RTFEYIAGCI
    180 301  EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQQQEVQQS
    210 351  AQQHSRAQV ATLGGKPKQN LMYDNWEERG KPFEIGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 321):

```

      1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
     40  51  AAAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
     80 101  ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAAATACCG
    120 151  CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
    160 201  GCAGCTTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
    200 251  TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
    240 301  TCGGCAGGTT CAAAAATCCC TGAATATGTC CAATGGCTGA ATACGCACAG
    280 351  ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCCT AAGCTTCTAG
    320 401  ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACAT CGCTTCAAAC
    360 451  AAGATGGGTA TCGGTACGCT TTTAGAAATGG AAAATATGCG CGGACGATCC
    400 501  CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
    440 551  AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
    480 601  AAGCGGTCAA AGTGGTTTTC CACTCTGCCA GTAATAGTAT TGCTGATTCC
    520 651  CGTGTGTTGTC GGCCTGTCCT ATAAAATGTT GAGCAGTTAC GGAAAAAAC
    560 701  AGGAAGAACC CGCAGCACA GAATCGGCGG CAACAGAACA GCAGGCAGTA

```

5

751	CTTCCGGATA	AAACAGAAAG	CGAGCCGGTA	AATAACGGCA	ACCTTACCGC
801	AGATATGTTT	GTTCCGACAT	TGTCCGAAAA	ACCCGAAAGC	AAGCCGATTT
851	ATAACGGTGT	AAGGCAGGTA	AGAACCTTTG	AATATATAGC	AGGCTGTATA
901	GAAGGCGGAA	GAACCGGATG	CGCCTGCTAT	TCGCATCAAG	GGACGGCATT
951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
1001	CGTTTAACCC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
1051	GCGCAGCAAC	ATTCGGACAG	GGCGCAAGTT	GCCACATTGG	GCGGAAAACC
1101	GTAGCAGAAC	CTAATGTACG	ATAATTGGGA	AGAACCGGGG	AAACCGTTTG
1151	AAGGAATCGG	CGGGGGCGTG	GTCGGATCGG	CAAACCTGA	

This corresponds to the amino acid sequence (SEQ ID NO: 322; ORF84-1):

	1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGIRRVK	FTNIKGLKIP
	51	HTYIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENIGSIVIV	DEAQDVWPAR
15	101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLLDQNLRTL	VRKHYHIASN
	151	KMGMRITLLEW	KICADDPVKM	ASSAFSSIYT	LDKKVVDLYE	SAEVHTVNVK
	201	KRSKWFYTLF	VIVLLIPVVF	<u>GLSYKMLSSY</u>	GKKQEPEAAQ	ESAATEQQAV
	251	LPDKTEGEPV	NNGNLTADMF	VPTLSEKPS	KPIYNGVRVQ	RTFEYIAGCI
	301	EGGRTGCACV	SHQGTALKEV	TELMCKDYPK	NGLPFPNYKE	ESQQQEVQQS
	351	AQQHSDRAQV	ATLGGKP*QN	LMYDNWEERG	KPFEGIGGGV	VGSAN*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 (SEQ ID NO: 320) shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) (SEQ ID NO: 324) from strain A of *N. meningitidis*:

25		10	20	30	40	50	60
	orf84.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDEKAI	RRKVFTNI	KGLKIPHTYIETDAK			
	orf84a	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENG	IRRVKVFTNI	KGLKIPHTYIETDAK			
		10	20	30	40	50	60
30		70	80	90	100	110	120
	orf84.pep	LPKSTDEQLSAHDMYEWIKK	PENIGSIVIVDEAQD	VWPARSAGSKIPENVQWLN	THR	HQG	
	orf84a	LPKSTDEQLSAHDMYEWIKK	PENIGSIVIVDEAQD	VWPARSAGSKIPENVQWLN	THR	HQG	
		70	80	90	100	110	120
35		130	140	150	160	170	180
	orf84.pep	IDIFVLTQGP	KLDDQNLR	TLVRKHYH	IASNKMGM	RTLLEWKICADDPVKMASSAFSSIYT	
	orf84a	IDIFVLTQGS	KLDDQNLR	TLVRKHYH	IASNKMGM	RTLLEWKICADDPVKMASSAFSSIYT	
		130	140	150	160	170	180
40		190	200	210	220	230	240
	orf84.pep	LDKKVYDLYXXAEVHTV	NKVKRSKW	FYTL	PVIVLLIPVFVGLSYKMLSSYGKKQEEPA	AAQ	
	orf84a	LDKKVYDLYESA	EVHTV	NKVKRSKW	FYTL	PVIVLLIPVFVGLSYKMLSSYGKKQEEPA	AAQ
		190	200	210	220	230	240
45		250	260	270	280	290	300
	orf84.pep	ESAATEQQAVLPDKTEGE	PPVNNGNLTADMFVPTLSEK	PXSKPIYNGVRQVRTFEYIAGCI			

	orf84a	ESAATEHQAVFQDKTEGEPVNNGNLTADMVFPVPTLSEKPESKPIYNGVRQVRTFEYIAGCV	250	260	270	280	290	300
			310	320	330	340	350	360
5	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHS						
	orf84a	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQSQEQHHS						
			310	320	330	340	350	360
			370	380	390			
10	orf84.pep	ATLGGKXPQNLMYDNWEERGKPFEGIGGGVVG						
	orf84a	ATLGGKPWQNLMYDNWQERGKPFEGIGGGVVG						
			370	380	390			

The complete length ORF84a nucleotide sequence (SEQ ID NO: 323) is:

15	1	ATGGCAGAGA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTCAG	GGAAACATT
	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAG	CCGGATGAAA
	101	ACGGCATACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGCTT	GAAGATACCG
	151	CACACCTACA	TAGAAACGGA	CGCGAAAAAG	CTGCCGAAAT	CGACAGATGA
	201	GCAGCTTTTC	GCGCATGATA	TGTACGAATG	GATAAAGAAG	CCCCGAAATA
20	251	TCGGGTCTAT	TGTCATTGTA	GATGAAGCTC	AAGACGTATG	CGCGCACGCG
	301	TCGGCAGGTT	CAAAAAATCCC	TGAAAAATGTC	CAATGGCTGA	ATACGCACAG
	351	ACATCAGGGC	ATTGATATAT	TTGTTTTGAC	TCAAGGCTCT	AAGTCTCTAG
	401	ATCAAAATCT	TAGAACGCTT	GTACGGAAAC	ATTACCACAT	CGCTTCAAAC
	451	AAGATGGGTA	TGCGTACGCT	TTTAGAATGG	AAAAATATGCG	CGGACGATCC
25	501	CGTAAAAATG	GCATCAAGCG	CATTCTCCAG	TATCTATACA	CTGGATAAAA
	551	AAGTTTATGA	CTTGATCGAA	TCAGCGGAAG	TTTATACCGT	AAATAAGGTC
	601	AAGCGGTCAA	AATGGTTTTA	TACTCTGCCA	GTAATAATAT	TGCTGATTCC
	651	CGTTTTTGTG	GGCCTGTCCT	ATAAAATGTT	AAGTAGTTAT	GGAAAAAAC
	701	AGGAAGAAC	CGCAGACAA	GAATCGGCGG	CAACAGAACA	TCAGGCAGTA
30	751	TTTCAGGATA	AAACAGAAGG	CGAGCCGGTA	AACAAACGTA	ACCTTACCGG
	801	AGATATGTTT	GTTCCGACAT	TGTCGGAAAA	ACCCGAAAGC	AAGCGGATTT
	851	ATAACGGTGT	AAGGCAGGTA	AGAACCTTTG	AATATATAGC	AGGCTGTGTA
	901	GAAGGCGGAA	GAACCGGATG	CACATGCTAT	TCGCATCAAG	GGACGGCATT
	951	GAAAGAAATT	ACAAAGGAAA	TGTGCAAGGA	TTACGCAAGA	AACGGATTGC
35	1001	CGTTTAAACC	ATATAAGAA	GAAAGCCAAG	GGCGGGATGT	CCAGCAAAGT
	1051	GAGCAGCAAC	ATTCGGACAG	ACCGCAAGTT	GCCACGTTGG	GCGGAAAGCG
	1101	GTGGCAAAAT	CTTATGTATG	ATAATTGGCA	GGAGCGCGGA	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGGCGTG	GTCGGATCGG	CAAACCTGA	

This encodes a protein having amino acid sequence (SEQ ID NO: 324):

40	1	MAEICLTGT	PGSGKTLKMV	SMMANDEMFK	PDENGIRRVK	FTNIKGLKIP
	51	HTYIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENIGSIVIV	DEAQDVWPAR
	101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGS	KLLDQNLRTL	VRKHYHIASN
	151	KMGMRITLEW	KICADDPVKM	ASSAFSSIYT	LDKKVVDLYE	SAEVHTVNVK
	201	KRSKWFTYTL	<u>VIILLIPVFV</u>	<u>GLSYKMLSSY</u>	GKKQEPEAAQ	ESAATEHQAV
45	251	FQDKTEGEPV	NNGNLTADMF	VPTLSEKPES	KPIYFNGVRQ	RTFEYIAGCV
	301	EGGRGTCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPNFPYKE	ESQGRDVQQS
	351	EOHHSRDRPO	ATLGKGPWON	LMYDNWQERG	KPFEGTGGGV	VGSAN*

ORF84a (SEQ ID NO: 324) and ORF84-1 (SEQ ID NO: 322) show 95.2% identity in 395 aa  
50 overlap:

-274-

		10	20	30	40	50	60
	orf84a.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK 					
5	orf84-1	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK 					
		10	20	30	40	50	60
	orf84a.pep	70	80	90	100	110	120
	orf84-1	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG 					
10		70	80	90	100	110	120
	orf84a.pep	130	140	150	160	170	180
	orf84-1	IDIFVLTQGSKLLDQNLRTLVRKHYHIASNKMGMRLLLEWKICADDPVKMASSAFSSIYT 					
15		130	140	150	160	170	180
	orf84a.pep	190	200	210	220	230	240
	orf84-1	LDKKVYDLYESAEVHTVNVKVRKSKWFYTLPIVILLIPVFGVLSYKMLSSYGKKQEEPAAQ 					
20		190	200	210	220	230	240
	orf84a.pep	250	260	270	280	290	300
	orf84-1	ESAATEHQAVFQDKTEGEPVNNGLTADMVFVPTLSEKPESKPIYNGVRQVRTFEYIAGCV 					
25		250	260	270	280	290	300
	orf84a.pep	310	320	330	340	350	360
	orf84-1	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQQSEQHHSRDPQV 					
30		310	320	330	340	350	360
	orf84a.pep	370	380	390			
	orf84-1	ATLGGKPXQNLMYDNWQERGKPFEGIGGGVVG SANX 					
35		370	380	390			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 (SEQ ID NO: 320) shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) (SEQ ID NO: 326) from *N. gonorrhoeae*:

40	orf84.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGVRKVFTNIKGLKIPHTIETDAKK	60
	orf84.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEWIKKPENVGAIIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120



	orf84.pep	IDIFVLTQGP <del>LLDQNLRTL</del> VKRHYHIA <del>SNKMG</del> MRTLLEWKICADDPVKMASSAFSSIIYT	180
	orf84ng	IDIFVLTQGP <del>LLDQNLRTL</del> VKRHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIIYT	180
5	orf84.pep	LDKKVYDLYXXAEVHTVNKVKRSKW <del>FYTL</del> PVIVLLIPV <del>FVGL</del> SYKMLSSYGKKQEEPAAQ	240
	orf84ng	LDKKVYDLYESAEIHTVNKVKRSKW <del>FYALP</del> VIILLIPLFVGLSYKMLGSYGKKQEEPAAQ	240
	orf84.pep	ESAA <del>TEQ</del> QAVLPDKTEGEPVNNGNLTADMFVPTLSEK <del>XP</del> SKPIYNGVRQV <del>RTFEY</del> IAGCI	300
	orf84ng	ESAA <del>TEQ</del> QAVLPDKTEGESVNNGNLTADMFVPTLPEK <del>PES</del> KPIYNGVRQV <del>RTFEY</del> IAGCI	300
10	orf84.pep	EGGRTGCAC <del>YSHQ</del> TALKEVTELMCKDYVKNGLPFNPYKEESQ <del>QEV</del> QSSAQ <del>QHS</del> DRAQV	360
	orf84ng	EGGRTGCTCYSHQ <del>TALKEV</del> TELMCKDYVKNGLPFNPYKEESQ <del>QEV</del> QSSAQ <del>QHS</del> DRAQV	360
15	orf84.pep	ATLGGK <del>PXQ</del> NLMYDNWEERGKPFEGIGGGV <del>VSAN</del>	395
	orf84ng	ATLGGK <del>PQ</del> QNLMYDNWEERGKPFEGIGGGV <del>VSAN</del>	395

The complete length ORF84ng nucleotide sequence (SEQ ID NO: 325) is:

	1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTCAG	GGAAAACATT
	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAG	CCAGATGAAA
20	101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG
	151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA
	201	ACAGCTTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
	251	tcggcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC
	301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
25	351	GCATCAGGGC	ATAGATATAT	TTGTATTGAC	ACAAGGTCTT	AAACTCTTAG
	401	ATCAGAACTT	GCGAACATTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
	451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
	501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
	551	AAGTTTATGA	CTTGTACGAA	TCCGCAGAAA	TTCACACGGT	AAACAAAGTC
30	601	AAGCGTTCAA	AATGGTTTTA	TGCATTGCC	GTCATCATAT	TATTGATTCC
	651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGCGCG	CAACAGAACA	GCAGGCAGTA
	751	CTTCCGGATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
	801	AGATATGTTT	GTTCCGACAT	TGCCCCGAAA	ACCCGAAAGC	AAGCCGATTT
35	851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
	901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
	951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
	1001	CGTTTAAACC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
	1051	GCGCAGCAAC	ATTCCGACAG	GGCGCAAGTT	GCCACCTTGG	GCGGAAAACC
40	1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACCGGGG	AAACCGTTTG
	1151	AAGGAATCGG	CGGGGGCGTG	GTCGGATCGG	CAAACTGA	

This encodes a protein having amino acid sequence (SEQ ID NO: 326):

	1	MAEICLIT <del>GT</del>	PGSGKTLK <del>MT</del>	SMMA <del>NDEM</del> FK	PDENGVR <del>RR</del> KV	FTNIKGLKIP
	51	HTHIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENVGAIVIV	DEAQDVWPAR
45	101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLLDQNLRTL	VKRHYHIAAN
	151	KMGLRTLLEW	KVCADDPVKM	ASSAFSSIIYT	LDKKVYDLYE	SAEIHTVNKV
	201	KRSKW <del>FYALP</del>	<del>VIILLIPLFV</del>	GLSYKMLGSY	GKKQEEPAAQ	ESAATEQ <del>QAV</del>
	251	LPDKTEGESV	NNGNLTADMF	VPTLPEK <del>PES</del>	KPIYNGVRQV	RTFEYIAGCI
	301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLPFNPYKE	ESQGQEVQSS
50	351	AQQHSDRAQV	ATLGGK <del>PQ</del> QN	LMYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng (SEQ ID NO: 326) and ORF84-1 (SEQ ID NO: 322) show 95.4% identity in 395 aa overlap:

5	orf84-1.pep	MAEICLITGTPGSGKTLKMSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK
	orf84ng	MAEICLITGTPGSGKTLKMSMMANDEMFKPDENGVRKVFNTIKGLKIPHTIETDAKK
		:     :
10	orf84-1.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG
	orf84ng	LPKSTDEQLSAHDMYEWIKKPENVGAIIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG
		:     :
15	orf84-1.pep	IDIFVLTQGPVKLLDQNLRLTLVRKHYHIAASNMGMRTLLEWKICADDPVKMASSAFSSIYT
	orf84ng	IDIFVLTQGPVKLLDQNLRLTLVRKHYHIAANKMGLRTLLEWKVCADDPVKMASSAFSSIYT
		:     :     :     :
20	orf84-1.pep	LDKKVVDLYESAIEHTVNVKVRKSWFYTLPIVIVLLIPVFGVLSYKMLSSYGKKQEEPAQAQ
	orf84ng	LDKKVVDLYESAIEHTVNVKVRKSWFYALPVIILLIPVFGVLSYKMLGSYGKKQEEPAQAQ
		:     :     :     :
25	orf84-1.pep	ESAATEQQAVLPDKTEGEPVNNGLTADMVPTLSEKPEKPIYNGVRQVRTFEYIAGCI
	orf84ng	ESAATEQQAVLPDKTEGESVNNGLTADMVPTLPEKPEKPIYNGVRQVRTFEYIAGCI
		:     :     :     :
30	orf84-1.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSRAQV
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSRAQV
		:     :     :     :
35	orf84-1.pep	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVVGSAAX
	orf84ng	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVVGSAAX
		:     :

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 39

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 327):

```

1  GTGGTTTTC TGAATGCCGA CAACGGGATA TTGGTTCAGG ACTTGCCTTT
51 TGAAGTCAAA CTGAAAAAAT TCCATATCGA TTTTACAAAT ACGGGTATGC
101 CGCGTGATTT CGCCAGCGAT ATTGAAGTGA CGGACAAGGC AACCGGTGAG
5  151 AAACTCGAGC GCACCATCCG CGTGAACCAT CCTTTGACCT TGCACGGCAT
201 CACGATTTAT CAGGCGAGTT TTGCCGACGG CGGTTCGGAT TTGACATTCA
251 AGGCGTGGA TTTGGGTGAT GCTTCGCGCG AGCCTGTCTG GTTGAAGGCA
301 ACATCCATAC ACCAGTTTCC GTTGGAAATT GGCAAACACA AATATCGTCT
351 TGAGTTCGAT CAGTTCACCT CTATGAATGT GGAGGACATG AGCGAGGGCG
10 401 CGGAACGGGA AAAAGCCTG AAATCCACGC TGCCCGATGT CCGCGCCGTT
451 ACTCAGGAAG GTCACAAATA CACCAAT... ..TACCG
501 TATCCGTGAT GCGCCAGGCC AGGCGGTCTGA ATATAAAAC TATATGCTGC
551 CGGTTTTGCA GGAACAGGAT TATTTTGGGA TTACCGGCAC GCGCAGCGC.
601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
15 651 AGCGGACACC TTTATGGCAT TGCGTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751 GAACAATTCA TGCTGGCTGC GGAAAACACG CTGAACATCT TTGCACAAAA
801 AGGCTATTTG GGATTGGACG AATTATTAC GTCCAATATC CCGAAAGAGC
851 AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CGGCGTGATG
20 901 AACGTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT
25 1101 TTTGGTCTAT CTC...

```

This corresponds to the amino acid sequence (SEQ ID NO: 328; ORF88):

```

1  MVFLNADNGI LVQDLPFVEK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITYI QASFADGGSD LTFKAWNLDG ASREPVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRAV
30 151 TQEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRXVAD ATKGAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKEQQDKMQG YFYEMLYGVM
301 NAALDETSTR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
35 351 SEVRSSGLQM TRSXGPLLVY L...

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 329):

```

1  ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCGC
51 TTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
40 101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTTCTGG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
45 401 AACGTATCTT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCGCTG GCGGGTTGA
551 TAGACAGTAA CCTGCTGTTG AAACGGGTA TGCTGACCGG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
50 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCGBA TGTGGTTTTC CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC

```

5  
 10  
 15  
 20  
 25

```

901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTTCGGA
951 TTTGACATTG AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCTG
1001 TGTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAATAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCTG AATATAAAAA
1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC
1801 CCGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GGCGTGGGTA TTGTTTTAG
1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGCAGCGA ACGGGATTTG
1951 CAGAAGGAAT TTCAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA
  
```

25 This corresponds to the amino acid sequence (SEQ ID NO: 330; ORF88-1):

30  
 35  
 40

```

1 MSKSRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ GFQGKTINRE
151 DGSVLIAAKK GTMNKWGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITIYQAS FADGSDLTG KAWNLDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVNMNA LDETIRRYGL
551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFS DGKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*
  
```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF88 (SEQ ID NO: 328) shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) (SEQ ID NO: 332) from strain A of *N. meningitidis*:

45  
 50

```

orff88.pep          10          20          30
                   MVFLNADNGILVQDLPFEVKLKKFHIDFYN
orff88a             : |||||
                   AKDFKPESILGASNLSFRGNVNISEGQSADVFLNADNGILVQDLPFEVKLKKFHIDFYN
                   210      220      230      240      250      260
                   40       50       60       70       80       90
  
```

	orf88.pep	TGMPRDFASDIEVTDKATGEKLER TIRVNHPLTLHGITIYQASFADGGSDLTFKAWN LGD
	orf88a	TGMPRDFASDIEVTDKATGEKLER TIRVNHPLTLHGITIYQASFADGGSDLTFKAWN LGD
		270 280 290 300 310 320
5	orf88.pep	ASREPVLKATS IHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLPDVR AV
	orf88a	ASREPVLKATS IHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLNDVR AV
		330 340 350 360 370 380
10	orf88.pep	TQEGHKYTNXXXXXX YRIRDAPGQAVEYKNYMLPVLQE QDYFWITGRSXLQQQYRWLR I
	orf88a	TQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYMLPVLQE QDYFWITGRSGLQQQYRWLR I
		390 400 410 420 430 440
15	orf88.pep	PLDKQLKADTFMALREF LKDGEGRKRXVADATKGAPAEIREQFMLAAENTLNIFAQKGYL
	orf88a	PLDKQLKADTFMALREF LKDGEGRKRLVADATKGAPAEIREQFMLAAENTLNIFAQKGYL
		450 460 470 480 490 500
20	orf88.pep	GLDEFITSNIPKEQQDK MQGYFYEMLYGVMNAALDET XTRYGLPEWQQDEARNRFL LHSM
	orf88a	GLDEFITSNIPKEQQDK MQGYFYEMLYGVMNAALDET IRRYGLPEWQQDEARNRFL LHSM
		510 520 530 540 550 560
25	orf88.pep	DAYTGLTEYPAPMLLQ LDGFSEVRSSGLQMTR SXGPLL VYL
	orf88a	DAYTGLTEYPAPMLLQ LDGFSEVRSSGLQMTRSPGALLVYL GSVLLVLGT VLMFYVREKR
		570 580 590 600 610 620
30	orf88a	AWVLFSDGKIRFAMSS ARSERDLQKEFPKHVESLQRLGKDLN HDX
		630 640 650 660 670

The complete length ORF88a nucleotide sequence (SEQ ID NO: 331) is:

35	1	ATGAGTAAAT	CCCGTAGATC	TCCCCCACTT	CTTTCCCGTC	CGTG GTTCGC
	51	TTTTTTCAGC	TCCATGCGCT	TTGCGGTCGC	TTTGCTCAGT	CTGCTGGGTA
40	101	TTGCATCGGT	TATCGGTACG	GTGTTGCAGC	AAAACCAGCC	GCAGACGGAT
	151	TATTTGGTCA	AATTCGGATC	GTTTTGGGCG	CAGATTTTGT	GTTTTCTGGG
45	201	ACTGTATGAC	GTCATGCTT	CGGCATGGTT	TGTCGTTATC	ATGATGTTT
	251	TGGTGGTTTC	TACCA GTTTG	TGCCTGATTC	GCAATGTGCC	GCCGTTCTGG
50	301	CGCGAAATGA	AGTCTTTTCG	GGAAAAGGTT	AAAGAAAAAT	CTCTGGCGGC
	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTGCGCCC	GAGGTTGCCA
	401	AACGTTATCT	GGAAGTACAA	GGTTTTCAGG	GAAAAACCAT	TAACCGTGAA
	451	GACGGGTCGG	TTCTGATTGC	CGCCAAAAAA	GGCACAATGA	ACAAATGGGG
	501	CTATATCTTT	GCCCATGTTG	CTTTGATTGT	CATTTGCCTG	GGCGGGTTGA
	551	TAGACAGTAA	CCTGCTGTTG	AAACTGGGTA	TGCTGACCGG	TCGGATTGTT
	601	CCGACAATC	AGGCGGTTTA	TGCCAAGGAT	TTCAAGCCCG	AAAGTATTTT
	651	GGGTGCGTCC	AATCTCTCAT	TAGGGGCAA	CGTCAATATT	TCCGAGGGGC
	701	AGAGTGCGGA	TGTGGTTTTT	CTGAATGCCG	ACAACGGGAT	ATTGGTTTAC
	751	GACTTGCCCT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTTACAA
	801	TACGGGTATG	CCGCGCGATT	TTGCCAGTGA	TATTGAAGTA	ACGGATAAGG
	851	CAACCGGTGA	GAAACTCGAG	CGCACCATCC	GCGTGAACCA	TCCTTTGACC
	901	TTGCACGGCA	TCACGATTTA	TCAGGCGAGT	TTTGCCGACG	GCGGTTCCGA

5  
10  
15  
20

```

951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTGC
1001 TGTTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTTACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTA AAAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TGCGGCAGGG CAGGCGGTCTG AATATAAAAA
1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GGCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCTGTC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTC
1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGCAGCGA ACGGGATTTG
1951 CAGAAGGAAT TTCCAACA CGTCGAGAGT CTGCAACGGC TCGCAAGGA
2001 CTTGAATCAT GACTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 332):

25  
30  
35

```

1  MSKSRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51  YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ FQGKTINRE
151 DGSVLIAAKK GTMNKWGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
201 PDNQAVYAKD FKPESILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITIYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLQQ QYRWLRIPLD
451 KQLKADTFMA LREFLDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
501 FAQKGYLGLD EFITSNIKE QQDKMQGYFY EMLYGMNAA LDETIRRYGL
551 PEWQQDEARN RFLLSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFSDGKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

```

40 ORF88a (SEQ ID NO: 332) and ORF88-1 (SEQ ID NO: 330) 100.0% identity in 671 aa overlap:

45  
50

```

orf88a.pep  MSKSRSPPLLSRPWFFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
|
orf88-1      MSKSRSPPLLSRPWFFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60

orf88a.pep  QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
|
orf88-1      QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120

orf88a.pep  SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL 180
|
orf88-1      SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL 180

orf88a.pep  GGLIDSNLLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240
|
orf88-1      GGLIDSNLLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240

```

	orf88a.pep	LNADNGILVQDLPFVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88-1	LNADNGILVQDLPFVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
5	orf88a.pep	LHGITIYQASFADGGSDLTFAKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88-1	LHGITIYQASFADGGSDLTFAKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88-1	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
10	orf88a.pep	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
15	orf88-1	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88-1	LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
20	orf88a.pep	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88-1	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD	672
	orf88-1	LQRLGKDLNHD	672

## 25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 (SEQ ID NO: 328) shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) (SEQ ID NO: 334) from *N. gonorrhoeae*:

	orf88.pep	MVFLNADNGILVQDLPFVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
30	orf88ng	MVFLNADNGMLVQDLPFVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88.pep	PLTLHGITIYQASFADGGSDLTFAKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFD	120
	orf88ng	PLTLHGITIYQASFADGGSDLTFAKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFD	120
35	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDRAVTQEGHKYTNXXXXXXXXYRIRDAPGQAVEYKN	180
	orf88ng	QFTSMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
	orf88.pep	YMLPVLQEQQDYFWITGTRSLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRXVAD	240
	orf88ng	YMLPILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVAD	240
40	orf88.pep	ATKGAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng	ATKDAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQDKMQGYFYEMLYGVM	300

orf88.pep	NAALDETXTRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
orf88ng	NAALDETIIRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
orf88.pep	TRSXGPLLVL	371
orf88ng	TRSPGALLVYLSVLLVLGTVFMFYVPPKRAWVLFNSXKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence (SEQ ID NO: 333) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 334):

10	1	MVFLNADNGM	LVQDLPPFEVK	LKKFHIDFYN	TGMPRDFASD	IEVTDKATGE
	51	KLERTIRVNH	PLTLHGITIY	QASFADGGSD	LTFKAWNLRD	ASREPVVLKA
	101	TSIHQFPLEI	GKHKYRLEFD	QFTSMNVEDM	SEGAEREKSL	KSTLNDVRVAV
	151	TQEGKKYTN	GPSIVYRIRD	AAGQAVEYKN	YMLPILQDKD	YFWLTGTRSG
	201	LQQQYRWLRI	PLDKQLKADT	FMALREFLKD	GEGRKRLVAD	ATKDAPAEIR
15	251	EQFMLAAENT	LNIFAQKGYL	GLDEFITSNI	PKGQQDKMQG	YFYEMLYGVM
	301	NAALDETIIR	YGLPEWQQDE	ARNRFLHSM	DAYTGLTEYP	APMLLQLDGF
	351	SEVRSSGLQM	TRSPGALLVY	LGSVLLVLGT	VFMFYVPPKR	AWVLFNSXKI
	401	RFAMSSARSE	RDLQKEFPKH	VESLQRLGKD	LNHD*	

20 Further work revealed the complete gonococcal DNA sequence (SEQ ID NO: 335):

	1	ATGAGTAAAT	CCCGTATATC	TCCCACACTT	CTTCCCGTC	CGTGGTTCGC
	51	TTTTTTCAGC	TCCATGCGCT	TTGCGGTCGC	TTTGCTCAGT	CTGCTGGGTA
	101	TTGCATCGGT	TATCGGCACG	GTGTTACAGC	AAAACCAGCC	GCAGACGGAT
	151	TATTTGGTCA	AATTCGGACC	GTTTGGACT	CGGATTTTGT	ATTTTTTGGG
25	201	TTTGATGAT	GTCTATGCTT	CGGCATGGTT	TGTCGTTATC	ATGATGTTTC
	251	TGGTGGTTTC	TACCAGTTTG	TGTTAATCC	GTAACGTTCC	GCCGTTTGG
	301	CGCGAAATGA	AGTCTTCCG	GGAAAAGGTT	AAAGAAAAT	CTCTGGCGGC
	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTGCCCCC	GAAGTTGCCA
	401	AACGTTATCT	GGAGGTGCGG	GGTTTTCAGG	GAAAAACCGT	CAGCCGTGAG
30	451	GACGGGTCGG	TTCTGATTGC	CGCCAAAAAA	GGCAaatga	acaaATGGGG
	501	CTATATCTTT	GCcCaagtag	ctTTGATTGT	CATTTGCCTG	GGCGGGTTGA
	551	TAGACAGTAA	CCTGCTGCTG	AAGCTGGGTA	TGCTGGCCGG	TCGGATTGTT
	601	CCGACAATC	AGGCGTTT	TGCCAAGGAT	TTCAAGCCCG	AAAGTATTTT
	651	GGGTGCGTCC	AATCTCTCAT	TAGGGGCAA	CGTCAATATT	TCCGAGGGGC
35	701	AAAGTGC	GAATGCGG	CTGAATGCCG	ACAACGGGAT	GTTGGTTCAG
	751	GACTTGCCTT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTTACAA
	801	TACGGGTATG	CCGCGCGATT	TTGCCAGCGA	TATTGAAGTA	ACGGACAAGG
	851	CAACCGGTGA	GAAACTCGAG	CGCACCATCC	GCGTGAACCA	TCCTTTGACC
	901	TTGCACGGCA	TCACGATTTA	TCAGGCGAGT	TTTGCCGACG	GCGGTTCCGA
40	951	TTTGACATT	CAAGCGTGGA	ATTGAGGGA	TGCTTCGCGC	GAACCTGTCTG
	1001	TGTTGAAGGC	AACCTCCATA	CACCAGTTTC	CGTTGGAAAT	CGGCAACAC
	1051	AAATATCGTC	TTGAGTTTCA	TCAGTTCACT	TCTATGAATG	TGGAGGACAT
	1101	GAGCGAGGGT	GCGGAACGGG	AAAAAAGCCT	GAAATCCACT	CTGAACGATG
	1151	TCCGCGCCGT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGGCCCTTCC
45	1201	ATCGTGTA	GCATCCGTGA	TGcggCAGGG	CAGGCGGTCTG	AATATAAAAA
	1251	CTATATGCTG	CCGATTTTGC	AGGACAAAGA	TTATTTTGG	CTGACCGGCA
	1301	CGCGCAGCGG	CTTGCAGCAG	CAATACCGCT	GGCTGCGTAT	CCCCTTGGAC
	1351	AAGCAGTTGA	AAGCGGACAC	CTTTATGGCA	TTGCGTGAGT	TTTTGAAAGA
	1401	TGGGGAAGGG	CGCAAACGTC	TGGTTGCCGA	CGCAACCAAA	GACGCACCTG
50	1451	CCGAAATCCG	CGAACAATTC	ATGCTGGCTG	CGGAAAACAC	GCTGAATATC
	1501	TTTGC	GCAAA	AAGGCTATTT	GGGATTGGAC	GAATTTATTA
	1551	CCCGAAAGGG	CAGCAGGATA	AGATGCAGGG	CTATTTCTAC	GAAATGCTTT
	1601	ACGGCGTGAT	GAACGCTGCT	TTGGATGAAA	CCATACGCCG	GTACGGCTTG
	1651	CCCGAATGGC	AGCAGGATGA	AGCGCGGAAC	CGTTTCCTGC	TGCACAGTAT



5  
 1701 GGATGCCTAT ACGGGGCTGA CGGAATATCC CGCGCCTATG CTGCTCCAGC  
 1751 TTGACGGGTT TTCCGAGGTG CGTTCCTCAG GTTTCAGAT GACCCGTTCG  
 1801 CCGGGTGCGC TTTTGGTCTA TCtcggctcg gtattggttg TTTTGGgtac  
 1851 ggtaTttatg tTTTATGTGC GCGAAAAACG GGCGTGGgta tTGTTTTcag  
 1901 aCGGCAAAAT CCGTTTTGCT ATGtCTTcgg CCcgacgca ACGGGATTTG  
 1951 cAGAaggaaT TTCCAAAACA CGtcgAGAGC CTGCAACggc tcggcaaggA  
 2001 CttgaaTCAT GACTga

This corresponds to the amino acid sequence (SEQ ID NO: 336; ORF88ng-1):

10  
 1 MSKSRLSPTL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQQNQPQTD  
 51 YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW  
 101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVR GFQGKTVSRE  
 151 DGSVLIAAKK GTMNKWGYIF AQVALIVICL GGLIDSNLLL KLGMLAGRIV  
 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGMLVQ  
 15 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT  
 301 LHGITYQAS FADGGSDLTF KAWNLRDASR EPVVLKATSI HQFPLEIGKH  
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS  
 401 IVYRIRDAAG QAVEYKNYML PILQDKDYFW LTGTRSGLQQ QYRWLRIPLD  
 451 KQLKADTFMA LREFLDGEG RKRLVADATK DAPAEIREQF MLAAENTLNI  
 20 FAQKGYLGLD EFITSNIPKG QQDKMQGYFY EMLYGVNNA LDETIRRYGL  
 551 PEWQQDEARN RFLHSDMAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS  
 601 PGALLVYLG S VLLVLGTVFM FYVREKRAWV LFS DGKIRFA MSSARSERDL  
 651 QKEFPKHVES LQRLGKDLNH D\*

25 ORF88ng-1 (SEQ ID NO: 336) and ORF88-1 (SEQ ID NO: 330) show 97.0% identity in 671 aa overlap:

orf88-1.pep MSKSRRSPPLLSRPWFAPFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGSFWA 60  
 ||||| || ||||||||||||||||||||||||||||||||||||||||| ||: 60  
 orf88ng-1 MSKSRLSPTLLSRPWFAPFSSMRFAVALLSLLGIASVIGTVLQQNQPQTDYLVKFGPFWT 60  
 30 orf88-1.pep QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120  
 :|| ||||||||||||||||||||||||||||||||||||||||| 120  
 orf88ng-1 RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120  
 35 orf88-1.pep SLLDVKIAPEVAKRYLEVQGFQGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL 180  
 |||||||||:|||||:|||||:|||||:|||||:|||||:|||||: 180  
 orf88ng-1 SLLDVKIAPEVAKRYLEVRGFQGKTVSREDGSVLIAAKKGTMNKWGYIFAQVALIVICL 180  
 orf88-1.pep GGLIDSNLLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240  
 |||||||||:|||||:|||||:|||||:|||||:|||||:||||| 240  
 orf88ng-1 GGLIDSNLLLKLGMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240  
 40 orf88-1.pep LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT 300  
 |||||||:|||||:|||||:|||||:|||||:|||||:||||| 300  
 orf88ng-1 LNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT 300  
 orf88-1.pep LHGITYQASFADGGSDLTFKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT 360  
 |||||||||:|||||:|||||:|||||:|||||:|||||:||||| 360  
 orf88ng-1 LHGITYQASFADGGSDLTFKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT 360  
 45 orf88-1.pep SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML 420  
 |||||||||:|||||:|||||:|||||:|||||:|||||:||||| 420  
 orf88ng-1 SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML 420

10

15

20

25

30

35

40

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 40**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 337):

```

5      1  ATGATGAGTA ATAmAATGGm ACAAAAAGGG TTTACATTGA TTGmGmTGAT
      51  GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
     101  ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
     151  GyCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
     201  CGATAATCAG ACCATCGAGA ACAAACCTGGA AATATTTGTC TCAGGCTATA
     251  AGATGAATCC GAAAATTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC
    10    301  GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGCGGTTT CGAAGGCGGG
     351  GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
     401  AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
     451  GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

```

15 This corresponds to the amino acid sequence (SEQ ID NO: 338; ORF89):

```

      1  MMSNXXMQKG FTLLXXMIVV AILGIISVIA IPSYXSYIEK GYQSOLYTEM
     51  XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNP KIA KKYSVSVKFV
    101  DKEKSRAYRL VGVPKAGTGY TSLVWMNSVG DGYKCRDAAS AQAHLETLLS
    151  DVGCEAFSNR KK*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 339):

```

      1  ATGATGAGTA ATAAAATGGA ACAAAAAGGG TTTACATTGA TTGAGATGAT
     51  GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
    101  ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
    151  GTCGGTATCA ACAATATTTT CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
    201  CGATAATCAG ACCATCGAGA ACAAACCTGGA AATATTTGTC TCAGGCTATA
    251  AGATGAATCC GAAAATTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
    301  GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGCGGTTT CGAAGGCGGG
    351  GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
    401  AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
    451  GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 340; ORF89-1):

```

    35    1  MMSNKMEQKG FTLLIEMMIVV AILGIISVIA IPSYQSYIEK GYQSOLYTEM
      51  VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNP KIA KKYSVSVKFV
    101  DKEKSRAYRL VGVPKAGTGY TSLVWMNSVG DGYKCRDAAS AQAHLETLLS
    151  DVGCEAFSNR KK*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with Pile of *N. gonorrhoeae* (accession number Z69260) (SEQ ID NO: 1135).

ORF89 (SEQ ID NO: 338) and Pile protein (SEQ ID NO: 1135) show 30% aa identity in 120a overlap:

```

orf89  8  QKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQFILKNPL- 66
           QKGFTLI  MIV+AI+GI++ +A+P+Y  Y  +  S+      G  +  ++ L  +  +
Pile    5  QKGFTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGIW 64

orf89  67 -DDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGYTLVSW 125
           DN  +      +G  + KI  KY  SV      +      GV  K  G  LS+W
Pile    65 PKDNTS-----AGVASSDKIKGYVQSVTVAKGVVTAEMASTGVNKEIQGKLSLW 115

```

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF89 (SEQ ID NO: 338) shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) (SEQ ID NO: 342) from strain A of *N. meningitidis*:

```

10      10      20      30      40      50      60
orf89.pep  MMSNXMXQKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQF
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf89a     MMSNKMEQKGFTLIXXXXXXAIXXXXSVIXXXYSYIEKGYQSQLYTEMVGINNISKQX
           10      20      30      40      50      60

15      70      80      90      100     110     120
orf89.pep  ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf89a     ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKGTGY
           70      80      90      100     110     120

20      130     140     150     160
orf89.pep  TLSVWMNSVGDGYKCRDAASAQAHALETLSDDVGCEAFSNRKKX
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf89a     TLSVWMNSVGDGYKCRDAASARAHLETLSDDVGCEAFSNRKKX
           130     140     150     160

```

The complete length ORF89a nucleotide sequence (SEQ ID NO: 341) is:

```

1  ATGATGAGTA ATAAATGGA ACAAAAAGGG TTTACATTGA TTGNGANGNT
51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT
101 ATCNNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
30 151 GTCGGTATCA ACAATATTTT CAAACAGTNT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCAAGA GCAAACTGGA AATATTTGTC TCAGGCTATA
251 AGATGAATCC GAAAATTGCC GAAAAATATA ATGTTTCGGT GCATTTTGTC
301 AATGAGGAAA AACCNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG
351 GACGGGTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
35 401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGTCCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 342):

```

1  MMSNKMEQKG FTLIXXXXXX AIXXXXSVIX XXXYSYIEK GYQSQLYTEM
40 51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHFV
101 NEEKPRAYSL VGVPKGTGY TLSVWMNSVG DGYKCRDAAS ARAHALETLS
151 DVGCEAFSNR KK*

```

ORF89a (SEQ ID NO: 342) and ORF89-1 (SEQ ID NO: 340) show 83.3% identity in 162 aa overlap:

-287-

		10	20	30	40	50	60
	orf89a.pep	MMSNKMEQKGFTLIXXXXXXAIXXXXSVIXXXSYIEKGYQSQLYTEMVGINNISKQX					
5	orf89-1	MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF					
		10	20	30	40	50	60
	orf89a.pep	ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKTGTGY					
10	orf89-1	ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY					
		70	80	90	100	110	120
	orf89a.pep	TLSVWMNSVGDGYKCRDAASARAHLETLSSDVGCEAFSNRKKX					
15	orf89-1	TLSVWMNSVGDGYKCRDAASAQAHLETLSSDVGCEAFSNRKKX					
		130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF89 (SEQ ID NO: 338) shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) (SEQ ID NO: 344) from *N. gonorrhoeae*:

20	orf89	MMSNXMXQKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQF	60
	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF	60
	orf89	ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY	120
	orf89ng	ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY	120
25	orf89	TLSVWMNSVGDGYKCRDAASAQAHLETLSSDVGCEAFSNRKK	162
	orf89ng	TLSVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKK	162

The complete length ORF89ng nucleotide sequence (SEQ ID NO: 343) is:

30	1	aTGATGAGCA	ATAAAATGGA	ACAAAAAGGG	TTTACATTGA	TTGAGATGAT
	51	GATAGTTGTC	ACGATACTCG	GCATCATCAG	CGTCATTGCC	ATACCTTCTT
	101	ATCAGAGTTA	TATTGAAAAA	GGCTATCAGT	CCCAGCTTTA	TACGGAGATG
	151	GTCGGTATCA	ACAAATGTTCT	CAAAACAGTTT	ATTTTGAAAA	ATCCCCAGGA
	201	CGATAATGAT	ACCCTCAAGA	GCAAACGTAA	AATATTTGTC	TCAGGCTATA
35	251	AGATGAATCC	GAAAAAttgCC	AAAAAATATA	GTGTTTCGGt	aaggtttGTC
	301	gatGCGGAAA	AACCAAGGGC	ATACAGGTTG	GTCGCGTTT	CGAACGCGGG
	351	GACGGGTAT	ACTTTGTCGG	TATGGATGAA	CAGCGTGGGC	GACGGATACA
	401	AATGCCGTGA	TGCCACTTCT	GCCCAGGCCT	ATTCGGACAC	CTTGTCCGCA
40	451	GATAGCGGCT	GTGAAGCTTT	CTCTAATCGT	AAAAAATAG	

This encodes a protein having amino acid sequence (SEQ ID NO: 344):

	1	MMSNKMEQKG	FTLIEMMIVV	TILGIISVIA	IPSYQSYIEK	GYQSQLYTEM
	51	VGINNVLKQF	ILKNPQDDND	TLKSKLKIFV	SGYKMNPKIA	KKYSVSVRFV
	101	DAEKPRAYRL	VGVPNAGTGY	TLSVWMNSVG	DGYKCRDATS	AQAYSDTLSA
45	151	DSGCEAFSNR	KK*			

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng (SEQ ID NO: 344) and ORF89-1 (SEQ ID NO: 340) show 88.3% identity in 162 aa overlap:

5		10	20	30	40	50	60
	orf89-1.pep	MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF					
	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF					
		10	20	30	40	50	60
10		70	80	90	100	110	120
	orf89-1.pep	ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY					
	orf89ng	ILKNPQDDNDTLKSKLKIFVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY					
		70	80	90	100	110	120
15		130	140	150	160		
	orf89-1.pep	TLNVWMNSVGDGYKCRDAASAQAHALETLSDDVGCEAFSNRKKX					
	orf89ng	TLNVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSNRKKX					
		130	140	150	160		
20							

Based on this analysis, including the gonococcal motifs and the homology with the known Pile protein (SEQ ID NO: 1135), it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 25 ORF89-1 (SEQ ID NO: 340) (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that ORF89-1 (SEQ ID NO: 340) is a surface-exposed protein, and that it is a
- 30 useful immunogen.

### Example 41

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 345):

35	1	ATGAAAAAAT	CCTCCCTCAT	CAGCGCATTG	GGCATCGGTA	TTTGAGCAT
	51	CGGCATGGCA	TTTGCCGCCC	CTGCCGACGC	GGTAAGCCAA	ATCCGTCAA
	101	ACGCCACTCA	AGTATTGAGC	ATCTTAAAAA	ACGGCGATGC	CAACACCGCT

```

151 CGCCAAAAAG CCGAAGCCTA TCGGATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
251 AACAAGCGTT GGCCn.AGAA TTTCAACCC...

```

5 This corresponds to the amino acid sequence (SEQ ID NO: 346; ORF91):

```

1 MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51 RQKAEAYAIP YFDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 347):

```

10      1 ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51 CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAAGCCAA ATCCGTCAAA
     101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
     151 CGCCAAAAAG CCGAAGCCTA TCGGATTCCC TATTTCGATT TCCAACGTAT
     201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
     15      251 AACAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
     301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
     401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
     20      501 CGTGTACCGC AACCAATTCTG GCGAAATTAT CAAAGCGAAA GGCGTGGACG
     551 GACTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAAATA A

```

This corresponds to the amino acid sequence (SEQ ID NO: 348; ORF91-1):

```

25      1 MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
     51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
     101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
     151 GKYRTYNVAI EGASLVTVYR NQFGEI IKAK GVDGLIAELK AKNGGK*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 (SEQ ID NO: 346) shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) (SEQ ID NO: 350) from strain A of *N. meningitidis*:

```

35      10      20      30      40      50      60
orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91a    MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
      10      20      30      40      50      60

      70      80      90
orf91.pep YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf91a    YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN
      70      80      90      100      110      120

orf91a    KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
      130      140      150      160      170      180

```

The complete length ORF91a nucleotide sequence (SEQ ID NO: 349) is:

```

1  ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
5  51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAACCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
251 AACAAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
10 351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
501 CGTGATCCGC AACCAATTCT GCGAAATTAT CAAAGCGAAA GGCGTGGACG
15 551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 350):

```

1  MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
51  RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FTLLIRTYS
101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
20 151 GKYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGSK*

```

ORF91a (SEQ ID NO: 350) and ORF91-1 (SEQ ID NO: 348) show 98.0% identity in 196 aa overlap:

```

25      10      20      30      40      50      60
orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
orf91-1     MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
           10      20      30      40      50      60

30      70      80      90      100     110     120
orf91a.pep YFDFQRM TALAVGNPWRTASDAQKQALAKEFTLLIRTYSGTMLKLKNANVNVKDNPIVN
orf91-1     YFDFQRM TALAVGNPWRTASDAQKQALAKEFTLLIRTYSGTMLKLKNANVNVKDNPIVN
           70      80      90      100     110     120

35      130     140     150     160     170     180
orf91a.pep KGGKEIIVRAEVLGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNVQFGEI IKAK
orf91-1     KGGKEIIVRAEVLGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYRNVQFGEI IKAK
           130     140     150     160     170     180

40      190
orf91a.pep GVDGLIAELKAKNGSKX
orf91-1     GVDGLIAELKAKNGGKX
           190

```

Homology with a predicted ORF from *N.gonorrhoeae*



ORF91 (SEQ ID NO: 346) shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) (SEQ ID NO: 352) from *N. gonorrhoeae*:

```

5  orf91.pep      MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP      60
    orf91ng      VKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP      60

    orf91.pep      YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP                      93
    orf91ng      YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPIVN      120

```

10 The complete length ORF91ng nucleotide sequence (SEQ ID NO: 351) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 352):

```

15 1  VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
    51 RPKAEAYAVP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTL LIRTYSG
    101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
    151 GK YRTYNVAI EGTSLVTYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 353):

```

20 1  ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
    51 CGGCATGGCA TTTGCCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA
    101 ACGCCACACA GGTTTTGACC ATCCTCAAAA GCGGCGACGC GGCTTCTGCA
    151 CGCCCAAAAG CCGAAGCCTA TCGCGTTCCC TATTTCGATT TCCAACGTAT
    201 GACCGCATTG GCGGTCGGCA ACCCTGGCG TACCGCGTCC GACGCGCAAA
    251 AACAAAGCGT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
    301 GGCACGATGC TGAAATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC
    25 351 CATCGTCAAT AAGGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
    401 TCCCCGGTCA GAAGCCGTC AATATGGACT TTACCACTA CCAAAGCGGC
    451 GGCAAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
    501 CGTGTACCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
    551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAAATA A
    30

```

This corresponds to the amino acid sequence (SEQ ID NO: 354; ORF91ng-1):

```

35 1  MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
    51 RPKAEAYAVP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTL LIRTYSG
    101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
    151 GK YRTYNVAI EGTSLVTYR NQFGEI I KAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 (SEQ ID NO: 354) and ORF91-1 (SEQ ID NO: 348) show 92.3% identity in 196 aa overlap:

```

40 10      20      30      40      50      60
    orf91-1.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
    orf91ng-1   MKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP
    10      20      30      40      50      60

    70      80      90      100     110     120
    45 orf91-1.pep YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKLKNANVNVKDNPIVN

```

-292-

```

      |||
orf91ng-1  YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPIVN
              70          80          90          100          110          120

      130          140          150          160          170          180
5  orf91-1.pep  KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
      |||:|||||:|||||
orf91ng-1  KGGKEIIVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEI IKAK
              130          140          150          160          170          180

      190
10 orf91-1.pep  GVDGLIAELKAKNGGKX
      |:|||||
orf91ng-1  GIDGLIAELKAKNGGKX
              190

```

15 In addition, ORF91ng-1 (SEQ ID NO: 354) shows homology to a hypothetical *E.coli* protein (SEQ ID NO: 1136):

```

      sp|P45390|YRBC_ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC REGION
      PRECURSOR (F211) )gi|606130 (U18997) ORF_f211 [Escherichia coli] )gi|1789583
20  (AE000399) hypothetical 24.0 kD protein in murZ-rpoN intergenic region [Escherichia
      coli]Length = 211

```

Score = 70.6 bits (170), Expect = 6e-12  
Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

```

25 Query: 59 VPHYDFQRM TALAVGNPWRTASDAQKQALAKEFQTL LIRTYSGTMLKFKNATVNVKDNPI 118
      +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
      Sbjct: 65 LPYVQVKYAGALVLGQYYSATPAQREAYFAAFREY LKQAYGQALAMYHGQTYQIA--PE 122

      Query: 119 VNKGKKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
      G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
      Sbjct: 123 QPLGDKTIVPIRVTIIDPNRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

30 Query: 175 EIIKAKGIDGLIAELKA 191
      +++ KGIDGL A+LK+
      Sbjct: 183 TLLR TKGIDGLTAQLKS 199

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes,  
35 could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 42

The following DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 355):

```

40      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTC AAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
      101 TGATCAGCA TACCCTCATC TCAAAATACA GTTTTGnnn nnnnnnnnnn
      151 nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
      201 CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGCGAA

```

251 AAGTCATCGT CTTCCGGCAG CCCAAAGCCG GCACGCCGCT GATGGTCAAA  
 301 GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC  
 351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG  
 401 GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA  
 451 AAAGTGATAC AAAAAACCGT AGGCGAATAA

This corresponds to the amino acid sequence (SEQ ID NO: 356; ORF97):

1 MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX  
 51 XXXXAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK  
 101 DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE  
 151 KLIQKTVGE\*

Further work revealed the complete nucleotide sequence (SEQ ID NO: 357):

1 ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTC AAC  
 51 CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA  
 101 TGACCACGCA TACCTCACC TCAAAATACA GTTTTGACGA AACCGTCAGC  
 151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATTT TTGCCGTCAT  
 201 CGACCATCAG GAAGCCGCCG GCCGAAACGG CTTAACGATG CAGCCGGCAA  
 251 AAGTCATCGT CTTCCGGCAG CCCAAAGCCG GCACGCCGCT GATGGTCAAA  
 301 GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC  
 351 GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG  
 401 GCAGCCGCAT CGGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA  
 451 AAAGTGATAC AAAAAACCGT AGGCGAATAA

This corresponds to the amino acid sequence (SEQ ID NO: 358; ORF97-1):

1 MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS  
 51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK  
 101 DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE  
 151 KLIQKTVGE\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 (SEQ ID NO: 356) shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) (SEQ ID NO: 360) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf97.pep		MKHILPLIAA	SALCISTASA	HPASEPSTQN	ETAMITHTLISK	YSFGXXXXXX	XXXXXAIKSKG
						:	:
orf97a		MXHILPLXXA	SALCISTASX	HPASEPQTQ	NETAMTHTLTS	SKYSFDETVSR	LETAIKSKG
		10	20	30	40	50	60
		70	80	90	100	110	120
orf97.pep		MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTPLMVKD	PAFALQLPLR	VLTETDGK
orf97a		MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTPLMVKD	PAFALQLPLR	VXVTETDGK
		70	80	90	100	110	120

```

              130      140      150      160
orf97.pep    VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
              |||||
orf97a       VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
              130      140      150      160

```

The complete length ORF97a nucleotide sequence (SEQ ID NO: 359) is:

```

      1  ATGANACACA  TACTCCCCCT  GANTGNCGCA  TCCGCACTCT  GCATTTCAAC
     51  CGCTTCGGNN  CATCTGCCA  GCGAACCACA  AACCCAAAAC  GAAACCGCTA
    101  TGACCACGCA  TACCCTCACC  TCAAAATACA  GTTTTGACGA  AACCGTCAGC
    151  CGCCTTGAAA  CCGCCATAAA  AAGCAAAGGG  ATGGACATTT  TTGCCGTCAT
    201  CGACCATCAG  GAAGCCGCCC  GCCGAAACGG  CTTAACGATG  CAGCCGGCAA
    251  AAGTCATCGT  CTTCCGCACG  CCCAAAGCCG  GTACGCCGCT  GATGGTCAAA
    301  GACCCCGCCT  TCGCCCTGCA  ACTGCCCTG  CGCGTCNTCG  TTACCGAAAC
    351  GGACGCAAAA  GTACGCGCCG  CCTATACCGA  TACGCGCGCC  CTCATCGCCG
    401  GCAGCCGCAT  CGGTTTCGAC  GAAGTGGCAA  ACACTTTGGC  AAACGCCGAA
    451  AAACGTATAC  AAAAAACCAT  AGGCGAATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 360):

```

      1  MXHILPLXXA  SALCISTASX  HPASEPQTQN  ETAMTHTLT  SKYSFDETVS
     51  RLETAIKSKG  MDIFAVIDHQ  EAARRNGLTM  QPAKVIVFGT  PKAGTPLMVK
    101  DPAFALQLPL  RVXVTETDGK  VRAAYTDTRA  LIAGSRIGFD  EVANTLANAE
    151  KLIQKTIGE*

```

ORF97a (SEQ ID NO: 360) and ORF97-1 (SEQ ID NO: 358) show 95.6% identity in 159 aa overlap:

```

              10      20      30      40      50      60
orf97a.pep    MXHILPLXXASALCISTASXHPASEPQTQNETAMTHTLTSTKYSFDETVSRLETAIKSKG
              |||||
orf97-1       MKHILPLIAASALCISTASAHASEPSTQNETAMTHTLTSTKYSFDETVSRLETAIKSKG
              10      20      30      40      50      60

              70      80      90      100     110     120
orf97a.pep    MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK
              |||||
orf97-1       MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK
              70      80      90      100     110     120

              130     140     150     160
orf97a.pep    VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
              |||||
orf97-1       VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX
              130     140     150     160

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 (SEQ ID NO: 356) shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) (SEQ ID NO: 362) from *N. gonorrhoeae*:

```

orf97.pep    MKHILPLIAASALCISTASAHASEPSTQNETAMITHLTISKYSFGXXXXXXXXXAISKSG    60

```

10

15

25

20

30

35

40

45

orf97-1.pep      130      140      150      160  
 VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX  
 ||:|||||||:||||:|||||||  
 orf97ng-1      VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX  
 130      140      150      160

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 ORF97-1 (SEQ ID NO: 358) (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure  
 15 12D). These experiments confirm that ORF97-1 (SEQ ID NO: 358) is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1 (SEQ ID NO: 358).

### Example 43

20 The following DNA, believed to be complete, sequence was identified in *N.meningitidis* (SEQ ID NO: 365):

1 ATGGCTTTTA TTACGCGCTT ATTCAAAAGC AGTAAATGGC TGATTGTGCC  
 51 GCTGATGCTC CCCGCCTTTC AGAATGTGGC GCGCGAGGGG ATAGATGTGA  
 101 GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC  
 151 CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TCGCGCGGGg  
 201 CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTCCGCC CCGATAATCG  
 251 CTTCTTATCG GTTTAAATG GGGCAACTGA TTGGCGATGA CGACaATATT  
 301 GACTACAAAC TGAGTTTCCA TCCGCTGACc AaACGCTACC GCGTTACCgT  
 351 CGgCGCGTTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA  
 30 CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT  
 401 CCGGAAGCAG GGGAAACCAA GGCGGAAATC CGCCTGACGC TGTCCACTTC  
 451 AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACTGGC  
 501 ATTTGGATTC GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA  
 551

35 This corresponds to the amino acid sequence (SEQ ID NO: 366; ORF106):

1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS  
 51 RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFKL GQLIGDDNII  
 101 DYKLSFHPLT KRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG

5

10

15

20

25

30

35

40

45

		10	20	30	40	50	59	
orf106.pep		MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ						
		::    :: :						
orf106a		MAFITRLFKSIKQWLVLPLMSVLPDAAAEGIDVSRAEARIXDGGQLSXXSRFQTELPDQ						
		10	20	30	40	50	60	
	60	70	80	90	100	110	119	
orf106.pep	LQQALRRGVPLNFTLSWQLSAPIIASYRFXKLGLIGDDDNIDYKLSFHPLTKRYRVTVGA							
	:							
orf106a	LQXAXRGVXLNXTLXWQLSAPIIASYRFXLGLIGDDDXIDYKLSFHPLTNRYRVTVGA							
	70	80	90	100	110	120		
	120	130	140	150	160	170	179	
orf106.pep	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAIEIRLTSTSKLPKPFQINALT							
orf106a	FSTXYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAIEIRLTSTSKLPKPFQINALT							
	130	140	150	160	170	180		
	180	190	199					
orf106.pep	SQNWHLDSGWKPLNIIGNKX							
orf106a	SQNWHLDSGWKPLNIIGNKX							
	190	200						

Due to the K→N substitution at residue 111, the homology between ORF106a (SEQ ID NO: 370) and ORF106-1 (SEQ ID NO: 368) is 87.9% over the same 199 aa overlap.

The complete length ORF106a nucleotide sequence (SEQ ID NO: 369) is:

```

5      1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
      51  GCTGCCGATG CTTTCCGTTT TGCCGACGCG GCGGCGGAG GGGATAGATG
     101  TGAGCCGCGC CGAAGCGAGG ATAANCGACG GCGGGCAGCT TTCCATNAGN
     151  AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAANNNG CGNNGNGCCG
     201  GGGCGTGNCG CTCAACTNTA CCTTAAGNTG GCAGCTTTCC GCCCGATAA
     251  TCGCTTCTTA TCGGTTTNA TGGGGCAAC TGATTGGCGA TGACGACNAT
     301  ATTGACTACA AACTGAGTTT CCATCCGCTG ACCAACCGCT ACCGCGTTAC
     351  CGTCGGCGCG TTTTCGACAG ANTACGACAC CTTGGATGCG GCATGCGCG
     401  CGACCGGCGC GGTTGCCAAC TGGAAAGTCC TGAACAAAGG CGCGCTGTCC
     451  GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC
    15   501  TTCAAACTG CCAAGCCTT TTCAAATCAA TGCATTGACT TCTCAAACT
      551  GGCATTTGGA TTCGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 370):

```

20      1  MAFITRLFKS IKQWLVL LPM LSVLPDAAAE GIDVSRAEAR IXDGGQLSXX
      51  SRFQTELPDQ LQXAXRGVX LNXTLXWQLS APIIASYRFX LGQLIGDDDX
     101  IDYKLSFHPL TNRYRVTGA FSTXYDTLDA ALRATGAVAN WKVLNKGALS
     151  GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLD SGW KPLNIIGNK*

```

#### Homology with a predicted ORF from *N. gonorrhoeae*

ORF106 (SEQ ID NO: 366) shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) (SEQ ID NO: 372) from *N. gonorrhoeae*:

```

25      orf106.pep  MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ  59
      orf106ng    MAFITRLFKSIKQWLVL LPI LSVLPDAAAE GIAATRAEARITDGGRLSISSRFQTELPDQ  60
      orf106.pep  LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGQLIGDDDNIDYKLSFHPLTKRYRVTGVA  119
      orf106ng    LQQALRRGVPLNFTLSWQLSAPTIASYRFLKGQLIGDDDNIDYKLSFHPLTNRYRVTGVA  120
      orf106.pep  FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT  179
      orf106ng    FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT  180
    35      orf106.pep  SQNWHLD SGWKPLNIIGNK  198
      orf106ng    SQNWHLD SGWKPLNIIGNK  199

```

Due to the K→N substitution at residue 111, the homology between ORF106ng (SEQ ID NO: 372) and ORF106-1 (SEQ ID NO: 368) is 91.0% over the same 199 aa overlap.



The complete length ORF106ng nucleotide sequence (SEQ ID NO: 371) is:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
51  GTTGCCGATA CTCTCCGTTT TGCCGGACGC GGCGGCGGAG GGCATTGCCG
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC
151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTTCC GCCCGACAA
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAAT
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCCTG ACCGCGTTAC
351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG
401 CGACCGGCGC GGTTGCCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCTGA CGCTGTCCAC
501 TTCAAACTG CCAAGCCTT TCCAATCAA CGCATTGACT TCTCAAACT
551 GGCATTGGA TTCGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

15 This encodes a protein having amino acid sequence (SEQ ID NO: 372):

```

1  MAFITRLFKS IKQWLVLPI LSVLPDAAE GIAATRAEAR ITDGGRLSIS
51  SRFQTELPDQ LQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
101 IDYKLSFHPL TNRYRVTVGA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLD SGW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (SEQ ID NO: 368) (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 (SEQ ID NO: 368) is a surface-exposed protein, and that it is a useful immunogen.

#### Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 373):

```

1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGGCGGG GCTgACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACAcCT TGTTCAAAC CCTGTTCTTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCCGCC GTCCCTGCCG

```

5  
10  
15  
20

```

301 TCTGAAATCC TGTTTTCACT CGACGATGCC gCCGCCGGCa TCGGGCTGGT
351 GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCcCAAG
451 CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGCTCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTCGCCCGC CGTCCTGCAC CGGGGG.TGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCTG
751 ATGGGTATTT CGTTCGGCGG GCGGCGATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCCTCC
901 GCCCTCTGC.TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGCTGCC
951 GGAAACTAC GCCGCCGTCC GGTTTATCGT CGTATCGTGT ATG.TGCCCG
1001 CGCTGTTTTG CACGTGGCG GAAATCAGCG GCATCGGTTT GAACGTGCTT
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGCGGAGG CCGCC.GGCG
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC
1201 GAAAGCTCyT GCCGCTGTG GCAGCCGCTC AAACGCTGC CGCTTTATCT
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301 CGCCGGCAAA CTATCCCTG TTTGCCGGCG TATGGGCGGC ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
1401 GAAAAACAA GGTTCCTAT TATGA

```

25 This corresponds to the amino acid sequence (SEQ ID NO: 374; ORF10):

30  
35

```

1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLMEGRAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGXRYGIPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSIFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCXTGIFSP LASLLLPEY AAVRFIVVSC MXPPLFCTLA EISIGLNVV
351 RKTRPIALAT LGALANLLL LGLDRAVPAR PXGAAVACAA SFWLFFAFKT
401 ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAyla
451 GCILRHRKDL HKLPHYLLKKQ GFPL*

```

Further sequence analysis revealed the complete DNA sequence(SEQ ID NO: 375) to be:

40  
45  
50  
55

```

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGCGGG GCTGACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACACCT TGTTCAAAAC CCTGTTCCCTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC GCCGCCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCcCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCTGA CCGCCGTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTCGCCCGC CGTCCTGCAC CGGGGGCTGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCTG
751 ATGGGTATTT CGTTCGGCGG GCGGCGATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCCGCCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCCTCC
901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTTGCCTCCC TCCTGCTGCC
951 GGAAACTAC GCCGCCGTCC GGTTTATCGT CGTATCGTGT ATGCTGCCGC

```

5  
1001 CGCTGTTTTC CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC  
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA  
1101 CCTGCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGGCGCGG  
1151 CGGTTGCCTG TGCCGCCTCA TTCTGGCTGT TTTTTCCTT CAAGACCGAA  
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCGC TTTATCTGCA  
1251 CACATTGTC TGCCTGACCT CCTCGGCGGC CTACACCTGC TTCGCACGC  
1301 CGGCAACTA TCCCCTGTTT GCCGGCGTAT GGGCGGCATA TCTGGCAGGC  
1351 TGCATCTGC GCCACCGGAA AGATTGCAC AAAGTGTTC ATTATTGAA  
1401 AAAACAAGGT TTCCCATAT GA

This corresponds to the amino acid sequence (SEQ ID NO: 376; ORF10-1):

15  
1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV  
51 SVLCLGLDQA YVREYYATAD KDTLFTLFL PPLLSAAAIA ALLLSRPSLP  
101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLMEGRAL AFSSAQLVPK  
151 LAILLLLPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR  
201 HAPFSPAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS  
251 MGISFGGAAL LFQSIFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS  
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLA EISGIGLNVV  
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFAFKTE  
20 401 SSCRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTPANYPLF AGVWAAAYLAG  
451 CILRHRKDLH KLFHYLKKQG FPL\*

Computer analysis of this amino acid sequence gave the following results:

#### Prediction

25 ORF10-1 (SEQ ID NO: 376) is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM (SEQ ID NO: 1137) from *Streptococcus thermophilus* (accession number U40830).

30 ORF10 (SEQ ID NO: 374) shows homology with the epsM gene of *S. thermophilus*, which encodes a protein (SEQ ID NO: 1137) of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

35 Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLQSIFSTVW 270  
L Y +PL SS+ +W L ++ R F+ + G G+ ++ + +IF+ W  
Sbjct: 210 LYYALPLIPSSILWLLNASSRYFVLFLLGAGANGLLAVATKIPSIISIFNTIFTQAW 267

Identities = 15/57 (26%), Positives = 31/57 (54%)

Query: 7 LGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63

-302-

L + G++GS +L +++PL ++ + G L QT A L + ++ + + A +R  
 Sbjct: 12 LVFTIGNLGSKLLVFLVPLYTYAMTPQEYGMADLYQTTANLLLPLITMNVFDATLR 68

Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLL PENYAAVRFTVVSCMLPPLFYTLTEISGIGLNVVRKTRPIXXXXXXXXXX 366

+ P+ ++ +YA+ V ML LF + ++ G ++T+ +

Sbjct: 305 VLKPIVEKVVSSDYASSWQYVPFFMLSMLFSSFSDFFGTNYIAAKQTKGVFMTSIYGTIV 364

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF10 (SEQ ID NO: 374) shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) (SEQ ID NO: 378) from strain A of *N. meningitidis*:

10	orf10.pep	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	10	20	30	40	50	60
	orf10a	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	10	20	30	40	50	60
15	orf10.pep	YVREYYATADKDTLTKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	70	80	90	100	110	120
	orf10a	YVREYYAAADKDTLTKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	70	80	90	100	110	120
20	orf10.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLPLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
25	orf10.pep	NLAAAFLLFQNRCLKAVRHAPFSPAHLHRGXRYGIPIALSSIAYWGLASADRLFLKKY	190	200	210	220	230	240
	orf10a	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAYWGLASADRLFLKKY	190	200	210	220	230	240
30	orf10.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	250	260	270	280	290	300
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS	250	260	270	280	290	300
35	orf10.pep	ALCXTGIFSPLASLLL PENYAAVRFIIVVSCMXPPLFCTLAEISGIGLNVVRKTRPIALAT	310	320	330	340	350	360
	orf10a	ALCLTGIFSPLASLLL PENYAAVRFIIVVSCMLPPLFCTLVEISGIGLNVVRKTRPIALAT	310	320	330	340	350	360
40	orf10.pep	LGALAANLLLLGLDRAVPA-R-PXGAAVACAASFVLFVFAFKTESSCRLWQPLKRLPLYLHT	370	380	390	400	410	419
	orf10a	LGALAANLLLLGL--AVPSGGARGA-AVACAASFVLFVFAFKTESSCRLWQPLKRLPLYMHT	370	380	390	400	410	

orf10.pep 420 430 440 450 460 470  
 LFCLTSSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKDLHLKLFHYLKKQGFPLX  
 orf10a 420 430 440 450 460 470  
 LFCCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHLKLFHYLKKQGFPLX

The complete length ORF10a nucleotide sequence (SEQ ID NO: 377) is:

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC  
 51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCTGCCG  
 101 ACGACATCGG ACGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG  
 151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC  
 201 CGCCGCCGAC AAAGACACTT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC  
 251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC ATCCCTGCCG  
 301 TCTGAAATCC TGTTTTCGCT CGACGATGCC GCCGCCGGCA TCGGGCTGGT  
 351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG  
 401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGTCCAAG  
 451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC  
 501 GGCGAACACC GCCGTCCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG  
 551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGCTGAA GGCCGTCGGG  
 601 CGCGCACCGT TTTATCCGC CGTCCTGCAT CGCGGCTGC GCTACGGCAT  
 651 ACCGATCGCA CTAAGCAGCA TCGCCTATTG GGGCTGGCA TCCGCCGACC  
 701 GTTTGTTTCT GAAAAAATAT GCCCGCCTAG AACAGCTCGG CGTTTATTTCG  
 751 ATGGGTATTT CGTTCGGCGG AGCGGCATTA TTGTTCCAAA GCATCTTTTC  
 801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGCA AACGCCCCGC  
 851 CCGCCCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC  
 901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTCGCCTCCC TCCTGCTGCC  
 951 GGAAACTAC GCCGCCGTCC GGTATATCGT CGTATCGTGT ATGCTGCCTC  
 1001 CGCTGTTTTG CACGTGGTA GAAATCAGCG GCATCGGTTT GAACGTCGTC  
 1051 CGAAAAACAC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA  
 1101 CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGGCGGCGCG CGCGGCGCGG  
 1151 CGGTTGCCCTG TGCCGCCTCA TTTTGGCTGT TTTTGTGTTT CAAGACCGAA  
 1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAAA CGCCTGCCGC TTTATATGCA  
 1251 CACATTGTTT TGCCTGGCCT CCTCGGCGGC CTACACCTGC TTCGGCACTC  
 1301 CGGCAAACTA CCCCCTGTTT GCCGGCGTAT GGGCGGTATA TCTGGCAGGC  
 1351 TGCATCCTGC GCCACCGGAA AGATTGACAC AAAGTGTTC ATTATTGAA  
 1401 AAAACAAGGT TTCCATTAT GA

This encodes a protein having amino acid sequence (SEQ ID NO: 378):

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV  
 51 SVLCLGLDQA YVREYYAAD KDTLTKTLFL PPLLSAAAIA ALLLSRPSLP  
 101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLRMGRAL AFSSAQLVSK  
 151 LAIIIIPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR  
 201 RAPFSSAVLH RGLRYGIPIA LSSIAWGLA SADRLFLKKY AGLEQLGVYS  
 251 MGISFGGAAL LFQSIFSTVW TPYIFRAIEA NAPPARLSAT AESAAALLAS  
 301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLV EISGIGLNVV  
 351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFVFKTE  
 401 SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAVYLAG  
 451 CILRHRKDLH KLFHYLKKQG FPL\*

50 ORF10a (SEQ ID NO: 378) and ORF10-1 (SEQ ID NO: 376) show 95.4% identity in 475 aa overlap:

orf10-1.pep 10 20 30 40 50 60  
 MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA

	orf10a	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	10	20	30	40	50	60
5	orf10-1.pep	YVREYYATADKDTLFTKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFEE	70	80	90	100	110	120
	orf10a	YVREYYAAADKDTLFTKTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFEE	70	80	90	100	110	120
10	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLPLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
15	orf10-1.pep	NLAAAFLLFQNRCLKAVRHAPFSPAHLRGXRYGIPIALSSIAYWGLASADRLFLKKY	190	200	210	220	230	240
	orf10a	NLAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAYWGLASADRLFLKKY	190	200	210	220	230	240
20	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	250	260	270	280	290	300
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS	250	260	270	280	290	300
25	orf10-1.pep	ALCXTGIFSPLASLLLPENYAARFIVVSCMXPPLFCTLAEISGIGLNVRKTRPIALAT	310	320	330	340	350	360
	orf10a	ALCLTGIFSPLASLLLPENYAARFIVVSCMLPPLFCTLVEISGIGLNVRKTRPIALAT	310	320	330	340	350	360
30	orf10-1.pep	LGALAAANLLLLGLDRAVPAR-PXGAAVACAASFVWFFAFKTESSCRLWQPLKRLPLYLHT	370	380	390	400	410	419
	orf10a	LGALAAANLLLLGL--AVPSGGARGAACAASFVWFFVFKTESSCRLWQPLKRLPLYMHT	370	380	390	400	410	
35	orf10-1.pep	LFCLTSSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKDLHKLFHYLKKQGFPLX	420	430	440	450	460	470
	orf10a	LFCLASSAAYTCFGTPANYPLFAGVWAVYLAGCILRHRKDLHKLFHYLKKQGFPLX	420	430	440	450	460	470

### Homology with a predicted ORF from *N.gonorrhoeae*

- 40 ORF10 (SEQ ID NO: 374) shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) (SEQ ID NO: 380) from *N. gonorrhoeae*:

orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
orf10nm	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60

	orf10ng.pep	YVREYYAAADKDTLTKTLFLPPLLSAAIAALLLSRPSLPSEILFSLDDAAAGIGLVLFE	120
	orf10nm	YVREYYATADKDTLTKTLFLPPLLSAAIAALLLSRPSLPSEILFSLDDAAAGIGLVLFE	120
5	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLPLTVGLLHFPANTSVLTAVALA	180
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLAILLXPLTVGLLHFPANTAVLTAVALA	180
	orf10ng.pep	NLAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPLALSSLAYWGLASADRLFLKKY	240
	orf10nm	NLAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAYWGLASADRLFLKKY	240
10	orf10ng.pep	AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAAALLAS	300
	orf10nm	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	300
15	orf10ng.pep	ALCLTGIFSPLASLLLPENYAARVFTVVSCLPPLFYTLTEISGIGLNVVRKTRPIALAT	360
	orf10nm	ALCXTGIFSPLASLLLPENYAARFIVVSCMXPPLFCTLAIEISGIGLNVVRKTRPIALAT	360
	orf10ng.pep	LGALAANLLLLGL--AVPSGGTRGA AVACAASFWLFFVKTESSCRLWQPLKRLPLYMHT	
20	orf10nm	LGALAANLLLLGLDRAVPAR-PXGA AVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT	
	orf10ng.pep	LFCLASSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKNLHKLFHYLKKQGFPLX	
25	orf10nm	LFCLTSSAAYTCFGTPANYPLFAGVWAAAYLAGCILRHRKDLHKLFHYLKKQGFPLX	

The complete length ORF10ng nucleotide sequence (SEQ ID NO: 379) is:

	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
30	51	GGTTTTAGCC	GTCATCATCC	TGCCGCTGCT	GTCGTGGTAT	TTCcccgCCG
	101	ACGACATCGG	GCGCATCGTG	CTGATGCAGA	CGGCGGCGGG	ACTGACGGTG
	151	TCGGTATTGT	GCCTCGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
	201	CGCCGCCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCCCTG	CCGCCGCTGC
	251	TGTTTTCCGC	CGCGATAGCC	GCCCTGCTGC	TTTCCCGCCC	GTCCCTGCCG
35	301	TCTGAAATCC	TGTTTTCGCT	CGACGATGCC	GCCGCCGGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTTCTCTTA	CTGTTTTCG
	401	GTATGGAAGG	GCGCGCCCTT	GCCTTTTCGT	CCGCGCAACT	CGTGCCCAAA
	451	CTCGCCATTC	TGCTGCTGTT	GCCGCTGACG	GTCGGGCTGC	TGCACTTTCC
	501	GGCGAACACC	TCCGTCCTGA	CCGCCGTTTA	CGCGCTGGCA	AACCTTGCCG
40	551	CCGCCGCCTT	TTTGCTGTTT	CAAAACCGAT	GCCGTCTGAA	GGCCGTCCCG
	601	CGCGCGCCGT	TTTCGCCCGC	CGTCCTGCAC	CGGGGGCTGC	GCTACGGCAT
	651	ACCGCTCGCA	CTGAGCAGCC	TTGCCTATTG	GGGGCTGGCA	TCCGCCGACC
	701	GTTTGTTCCCT	GAAAAAATAT	GCGGGCCTGG	AACAGCTCGG	CGTTTATTTCG
	751	ATGGGTATTT	CGTTCGGCGG	GGCGGCATTA	TTGCTCCAAA	GCATCTTTTC
	801	AACGGTCTGG	ACACCGTATA	TTTTCCGTGC	AATCGAAGAA	AACGCCACGC
45	851	CCGCCCGCCT	CTCGGCAACG	GCAGAATCCG	CCGCCGCCCT	GCTTGCCTCC
	901	GCCCTCTGCC	TGACCGGAAT	TTTCTCGCCC	CTCGCCTCCC	TCCTGCTGCC
	951	GGAAACTAC	GCCGCCGTCC	GGTTACCGT	CGTATCGTGT	ATGCTGccgc
	1001	cgctGTTTTA	CACGCTGACC	GAAATCAGCG	GCATCGGTTT	GAACGTCGTC
	1051	CGCAAAACGC	GTCCGATCGC	GCTTGCCACC	TTGGGCGCGC	TGGCGGCAAA
50	1101	CCTGCTGCTG	CTGGGGCTTG	CCGTACCGTC	CGGCGGCACG	CGCGGCGCGG
	1151	CGGTTGCCTG	TGCCGCCTCA	TTCTGGTTGT	TTTTTGTTTT	CAAGACAGAA
	1201	AGCTCCTGCC	GCCTGTGGCA	GCCGCTCAAA	CGCCTGCCGC	TTTATATGCA

1251 CACATTGTTC TGCCTgGCCT CCTCGGCGGC CTACACCTGC TTCGGCACAC  
 1301 CGGCAAACTA CCCcctgttt gccggcgtAT GGGCGGCATA TCTGGCAGGC  
 1351 TGCATCCTGC GCCACCGGAA AAATTTGCAC AAAGTGTTC ATTATTGAA  
 1401 AAAACAAGGT TTCCATTAT GA

5

This encodes a protein having amino acid sequence (SEQ ID NO: 380):

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV  
 51 SVLCLGLDQA YVREYYAAAD KDTLFKTLFL PPLLFSAAIA ALLLSRPSLP  
 101 SEILFSLDDA AAGIGLVLF LSF LPIRFL LVL RMEGRAL AFSSAQLVPK  
 151 LAI LLLPLT VGL LHFPANT SVLTAVYALA NLAAAFLLF QNRCRLKAVR  
 201 RAPFSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS  
 251 MGISFGGAAL LLQSIFSTVW TPYIFRAIEE NATPARLSAT AESAAALLAS  
 301 ALCLTGIFSP LASLLL PENY AAVRFTV VSC ML PPLFYTLT EISGIGLVV  
 351 RKTRPIALAT LGALANLLL LGLAVPSGGT RGA AVACAAS FWLFFVFKTE  
 15 401 SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAAAYLAG  
 451 CILRHRKNLH KLFHYLKQKQ FPL\*

ORF10ng (SEQ ID NO: 380) and ORF10-1 (SEQ ID NO: 376) show 96.4% identity in 473 aa overlap:

20 orf10-1.pep 10 20 30 40 50 60  
 orf10ng-1 MDTKEILGYAAGSIGSAVLA VIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA  
 25 orf10-1.pep 70 80 90 100 110 120  
 orf10ng-1 YVREYYATADKDTL FKTFLP PLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLF  
 30 orf10-1.pep 130 140 150 160 170 180  
 orf10ng-1 LSF LPIRFL LVL RMEGRALAFSSAQLVPKLA I LLLPLTVGL LHFPANTAVLTAVYALA  
 35 orf10-1.pep 190 200 210 220 230 240  
 orf10ng-1 NLAAAFLLFQNRCLKAVRHAPFSPAVLH RGLRYGIPIALSSIAYWGLASADRLFLKKY  
 40 orf10-1.pep 250 260 270 280 290 300  
 orf10ng-1 AGLEQLGVYSMGISFGGAALLQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS  
 45 orf10-1.pep 310 320 330 340 350 360  
 orf10ng-1 ALCLTGIFSP LASLLL PENY AAVRFTV VSC ML PPLFYTLT EISGIGLVV RKTRPIALAT



		370	380	390	400	410	420
orf10-1.pep	LGALAANLLLLGLAVPSGGARGAAVACAASF	FWLFFAFKTESSCRLWQPLKRLPLYLHTLF					
orf10ng-1	LGALAANLLLLGLAVPSGGTRGA	AVACAASF	FWLFFVFKTESSCRLWQPLKRLPLYMHTLF				
5		370	380	390	400	410	420
		430	440	450	460	470	
orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVWAA	YLAGCILRHRKDLHKL	FHYLKKQGFPLX				
orf10ng-1	CLASSAAYTCFGTPANYPLFAGVWAA	YLAGCILRHRKNLHKL	FHYLKKQGFPLX				
10		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 15 Example 45

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 381):

	1..	ATCCTGAAAC	CGCATAACCA	GCTTAAGGAA	GACATCCAAC	CTGATCCGGC
	51	CGATCAAAAC	GCCTTGTCCG	AACCGGATGC	TGCGACAGAG	GCAGAGCAGT
20	101	CGGATGCGGA	AAATGCTGCC	GACAAGCAGC	CCGTTGCCGA	TAAAGCCGAC
	151	GAGGTTGAAG	AAAAGGCGGG	CGAGCCGGAA	CGGGAAGAGC	CGGACGGACA
	201	GGCAGTGCGT	AAGAAAGCGC	TGACGGAAGA	GCGTGAACAA	ACCGTCAGGG
	251	AAAAAGCGCA	GAAGAAAGAT	GCCGAAACGG	TTAAAATACA	AGCGGTAAAA
	301	CCGTCTAAAG	AAACAGAGAA	AAAAGCTTCA	AAAGAAGAGA	AAAAGGCGGC
	351	GAAGGAAAAA	GTTGCACCCA	AACCAACCCC	GGAACAAATC	CTCAACAGCG
25	401	GCAgCATCGA	AAAmGCGCGC	AgTGCCGCCG	CCAAAGAAGT	GCAGAAAATG
	451	AA.AACGTCC	GACAAGGCGG	AAGC.AACGC	ATTATCTGCA	AATGGGCGCG
	501	TATGCCGACC	GTCAGAGCGC	GGAAGGGCAG	CGTGCCAAAC	TGGCAATCTT
	551	GGGCATATCT	TCCAAGGTGG	TCGGTTATCA	GGCGGGACAT	AAAACGCTTT
30	601	ACCGGGTGCA	AAGCGGCAAT	ATGTCTGCCG	ATGCGGTGA	

This corresponds to the amino acid sequence (SEQ ID NO: 382; ORF65):

	1..	ILKPHNQLKE	DIQDPADQN	ALSEPDAATE	AEQSDAENAA	DKQPVADKAD
	51	EVEEKAGEPE	REEPDGQAVR	KKALTEEREQ	TVREKAQKKD	AETVKIQAVK
35	101	PSKETEKKAS	KEEKKAAKEK	VAPKPTPEQI	LNSGSIEXAR	SAAAKEVQKM
	151	XNVRQGSXR	IICKWARMPT	VRARKGSVPN	WQSWAYLPRW	SVIRRDIKRF
	201	TGCKAAICLP	MR*			

Further work revealed the complete nucleotide sequence (SEQ ID NO: 383):

	1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGGTTTTTT
	51	CTTCGGTTTG	ATACTGGCGA	CGGTCATTAT	TGCCGGTATT	TTGTTTTATC
40	101	TGAACCAGAG	CGGTCAAAAT	GCGTTCAAAA	TCCCGGCTTC	GTCGAAGCAG
	151	CCTGCAGAAA	CGGAAATCCT	GAAACCGAAA	AACCAGCCTA	AGGAAGACAT
	201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGCTGCCA

5

10

15 This corresponds to the amino acid sequence (SEQ ID NO: 384; ORF65-1):

20

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF65 (SEQ ID NO: 382) shows 92.0% identity over a 150aa overlap with an ORF (ORF65a)  
(SEQ ID NO: 386) from strain A of *N. meningitidis*:

30

35

40

45

The complete length ORF65a nucleotide sequence (SEQ ID NO: 385) is:

```

1  ATGTTTATGA  ACAAATTTTC  CCAATCCGGA  AAAGGTCTGT  CCGGTTTTTT
51  CTTCGGTTTG  ATACTGGCGA  CGGTCATTAT  TGCCGGTATT  TTGTTTATC
101 TGAACCAGAG  CGGTCAAAAT  GCGTTCAAAA  TCCCAGTTCC  GTCGAAGCAG
5   151  CCTGCAGAAA  CGGAAATCCT  GAAACCGAAA  AACCAGCCTA  AGGAAGACAT
201 CCAACCTGAA  CCGGCCGATC  AAAACGCCTT  GTCCGAACCG  GATGCTGCGA
251 AAGAGGCAGA  GCAGTCGGAT  GCGGAAAAAG  CTGCCGACAA  GCAGCCCGTT
301 GCCGACAAAG  CCGACGAGGT  TGAGGAAAAG  GCGGACGAGC  CGGAGCGGGA
351 AAAGTCGGAC  GGACAGGCAG  TGCGCAAGAA  AGCACTGACG  GAAGAGCGTG
10  401  AACAAACCGT  CGGGGAAAAA  GCGCAGAAGA  AAGATGCCGA  AACGGTTAAA
451 AAACAAGCGG  TAAAACCATC  TAAAGAAACA  GAGAAAAAAG  CTTCAAAGA
501 AGAGAAAAAG  GCGGAGAAAG  AAAAAGTTGC  ACCCAAACCG  ACCCCGGAAC
551 AAATCCTCAA  CAGCGGCAGC  ATCGAAAAAG  CGCGCAGTGC  CGCTGCCAAA
601 GAAGTGCAAG  AAATGAAAAC  GCCCGACAAG  GCGGAAGCAA  CGCATTATCT
15  651  GCAAATGGGC  GCGTATGCCG  ACCGCCGAG  CGCGGAAGGG  CAGCGTGCCA
701 AACTGGCAAT  CTTGGGCATA  TCTTCCAAGG  TGGTCGGTTA  TCAGGCGGGA
751 CATAAACGCG  TTTACCGGGT  GCAAAGCGGC  AATATGTCTG  CCGATGCGGT
801 GAAAAAATG  CAGGACGAGT  TGAAAAACA  TGAAGTCGCC  AGCCTGATCC
20  851  GTTCTATCGA  AAGCAAATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 386):

```

1  MFMNKFSQSG  KGLSGFFFGL  ILATVIIAGI  LFYLNQSGQN  AFKIPVPSKQ
51  PAETEILKPK  NQPKEDIQPE  PADQNALSEP  DAAKEAEQSD  AEKAADKQPV
25  101  ADKADEVEEK  ADEPEREKSD  QQAVRKKALT  EEREQTVGEK  AQKKDAETVK
151  KQAVKPSKET  EKKASKEEKK  AEKEKVAPKP  TPEQILNSGS  IEKARSAAAK
201  EVQKMKTPDK  AEATHYLQMG  AYADRRSAEG  QRAKLAILGI  SSKVVGYQAG
251  HKTLYRVQSG  NMSADAVKKM  QDELKKHEVA  SLIRSIESK*

```

ORF65a (SEQ ID NO: 386) and ORF65-1 (SEQ ID NO: 384) show 96.5% identity in 289 aa overlap:

```

10      20      30      40      50      60
orf65a.pep  MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKIPVPSKQPAETEILKPK
35  orf65-1  MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK
10      20      30      40      50      60

70      80      90      100     110     120
orf65a.pep  NQPKEDIQPEPADQNALSEPDAAKEAEQSDAEKAADKQPVADKADEVEEKADEPEREKSD
40  orf65-1  NQPKEDIQPEPADQNALSEPDAAEAQSDAEKAADKQPVADKADEVEEKAGEPEREEDPD
70      80      90      100     110     120

130     140     150     160     170     180
orf65a.pep  QQAVRKKALTEEREQTVGEKAQKKDAETVKKQAVKPSKETEKKASKEEKKAEKEKVAPKP
45  orf65-1  QQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAAKEKVAPKP
130     140     150     160     170     180

190     200     210     220     230     240
orf65a.pep  TPEQILNSGSIEKARSAAAEVQKMKTPDKAEATHYLQMGAYADRRSAEGQRAKLAILGI
50  orf65-1  TPEQILNSGSIEKARSAAAEVQKMKTSKAEATHYLQMGAYADRQSAEGQRAKLAILGI
190     200     210     220     230     240

```

-310-

5

	250	260	270	280	290
orf65a.pep	SSKVVG YQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKY				
orf65-1	SSKVVG YQAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKY				
	250	260	270	280	290

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 (SEQ ID NO: 382) shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) (SEQ ID NO: 388) from *N. gonorrhoeae*:

10	ORF65ng	IIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLNQPKEDIQPEPADQNALSEP	30
	ORF65	ILKPHNQLKEDIQDPADQNALSEPDAATE	30
		10 20 30	
15	ORF65ng	AEQSDAEKAADKQPVADKADEVEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQKKD	90
	ORF65	AEQSDAENAADKQPVADKADEVEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQKKD	90
		40 50 60 70 80 90	
20	ORF65ng	AETVKKKAVKPSKETEEKKAAKEKVAPKPTPEQILNSRSIEKARSAAAEVQKM	150
	ORF65	AETVKIQAVKPSKETEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAEVQKM	150
		100 110 120 130 140 150	
25	ORF65ng	KNFGQGGSQRIICKWARMNPNGARKGSPVNWQSWAYLPKWSAIRRDIKRFTACKAAICPP	210
	ORF65	XNVRQGGSXRIICKWARMPTVRARKGSPVNWQSWAYLPRWSVIRRDIKRFTGCKAAICLP	210
		160 170 180 190 200 210	
30	ORF65ng	MR	
	ORF65	MR	

An ORF65ng nucleotide sequence (SEQ ID NO: 387) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 388):

35	1	MFMNKFSSQG	KGLSGFFFL	ILATVIIAGI	LLYLNQGGQN	AFKIPAPSKQ
	51	PAETEILKLK	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADKQPV
	101	ADKADEVEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKAETVK
	151	KKAVKPSKET	EKKASKEEK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
	201	EVQKMKNFGQ	GGSQRIICKW	ARMPNPGARK	GSVPNWQSWA	YLPKWSAIRR
40	251	DIKRFTACKA	AICPPMR*			

After further analysis, the complete gonococcal DNA sequence (SEQ ID NO: 389) was found to be:

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTCTT  
 51 CTTCGGTTTG ATACTGGCAA CGGTCAATAT TGCCGGTATT TTGCTTTATC  
 101 TGAACCAGGG CGGTCAAAAT GCGTTCAAAA TCCCGGCTCC GTCGAAGCAG  
 151 CCTGCAGAAA CGGAAATCCT GAAACTGAAA AACCAGCCTA AGGAAGACAT  
 5 201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGTTGCGA  
 251 AAGAGGCAGA GCAGTCCGAT GCGGAAAAAG CTGCCGACAA GCAGCCCCTT  
 301 GCCGACAAag cgcacgAGGT TGAAGAAAag GcGGgcgAgc cggAACGGga  
 351 aGAGCCGGAC ggACAGGCAG TGCGCAAGAA AGCACTGAcg gAAGAgcGTG  
 401 AACAAACcgt cagggAAAAA GCGCagaaga AAGATGCCGA AACGgTTAAa  
 10 451 AAacaaGCgg tAaaaccgtc tAAAGAAACa gaaaaaaaag cTtcaaaaga  
 501 agaaaaaaag gcggcgaaaag aaaAAGttgc acccaaaaccg accccggaaC  
 551 aaatcctcaa cagccgCagc atcgaaaaag cgcgtagtgc cgctgccaaa  
 601 gaAgtgcaGA AAatgaaaaa ctTtgggcaa ggcgGaagcc aacgcattaT  
 651 CTGcaaatgg gcgcgatatgc cgaccgtccg gagcgcggaA gggcagcgtg  
 15 701 ccaaACtggc aAtcttgGgc atatctTccg aagtggtcgG CTATCAGGCG  
 751 GGACATAAAA CGCTTTACCG CGTGCAaagc GGCAatatgt ccgccgatgc  
 801 gGTGAAAAAA ATGCAGGACG AGTTGAAAAA GCATGGGGtt gcCAGCCTGA  
 851 TCCGTGcgAT TGAAGGCAAA TAA

20 This encodes the following amino acid sequence (SEQ ID NO: 390):

1 MFMNKFSQSG KGLSGFFFL ILATVIIAGI LLYLNQGGQN AFKIPAPSKQ  
 51 PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV  
 101 ADKADEVEEK AGEPEREEDP GQAVRKKALT EEREQTVREK AQKKDAETVK  
 151 KQAVKPSKET EKKASKEEKK AAKEKVAPKP TPEQILNSRS IEKARSAAAK  
 25 201 EVQMKMNFQ GGSQRIICKW ARMPTVRS AE QRAKLAILG ISSEVVGYYA  
 251 GHKTLYRVQS GNMSADAVKK MQDELKKHGV ASLIRAIIEGK \*

ORF65ng-1 (SEQ ID NO: 390) and ORF65-1 (SEQ ID NO: 384) show 89.0% identity in 290 aa overlap:

30 orf65-1.pep 10 20 30 40 50 60  
 orf65ng-1 MFMNKFSQSGKGLSGFFFLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK  
 35 orf65-1.pep 70 80 90 100 110 120  
 orf65ng-1 NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP  
 40 orf65-1.pep 130 140 150 160 170 180  
 orf65ng-1 GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKASKEEKKAAKEKVAPKP  
 45 orf65-1.pep 190 200 210 220 230 239  
 orf65ng-1 TPEQILNSGSIEKARSAAAEVQMKMTSDKAEATHYL-QMGAYADRQSAEQRAKLAILG  
 50 orf65-1.pep 240 250 260 270 280 290  
 orf65ng-1 ISSKVVGYYAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX

orf65ng-1      |||:|||||||||||||||||||||||||||||||||:|:|  
 ISSEVVGYQAGHKTL YRVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX  
                   250                  260                  270                  280                  290

On this basis, including the presence of a putative transmembrane domain in the gonococcal  
 5 protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes,  
 could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID  
 NO: 391):

```

10      1  ATGAACCACG ACATCACTTT CCTCACCTG TTCCTACTCG GTkTCTTCGG
      51  CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GcGTTTGs.s
     101  TCCAAC TCCC CCCGCATATC AACCGCTTTT GGCTGATCCT GCTGCTTAAC
     151  ACAGGACGGG TAAGCAGCTA TACGGCAAtC GGCTGATAC TCGGATTAAT
     201  CGGACAGGTC GGCGTTTCAC TCGAcCAaAC CCGGTCCTG CAGAATATTT
     15  251  TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
     301  GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAaATCGGCA AACCGATATG
     351  GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
     401  CCGCCTGCCT tGCGgTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
     451  GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AgCGGTAGTG CGGCAACGGG
     20  501  CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTTtTAG
     551  CAATCGGCAT TTTtTCCCTG CAACTGAaAw AAATCATGCA AAACCGATAT
     601  ATCCGCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
     651  TGCCGTCCTG TGGCTGTAA
  
```

25 This corresponds to the amino acid sequence (SEQ ID NO: 392; ORF103):

```

      1  MNHDITFLTL FLLGXFGGTH CIGMCGGLSS AFXXQLPPhi NRFWLILLN
     51  TGRVSYTAI GLILGLIGQV GVSLDQTRVL QNILYTAANL LLLFLGLYLS
     101  GISSLAakIE KIGKPIWRNL NPILNRLlPI KSIPAClAVG ILWGWLPCGL
     151  VYSASLYALG SGSAATGGly MLAFALGTLP NLLAIGIFSL QLXKIMQnRY
     30  201  IRLCTGLSVS LWALWKLAVL WL*
  
```

Further work elaborated the DNA sequence (SEQ ID NO: 393) as:

```

      1  ATGAACCACG ACATCACTTT CCTCACCTG TTCCTACTCG GTTTCTTCGG
     51  CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGCGC
     35  101  TCCAAC TCCC CCCGCATATC AACCGCTTTT GGCTGATCCT GCTGCTTAAC
     151  ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCTGATAC TCGGATTAAT
     201  CGGACAGGTC GGCGTTTCAC TCGACCAAAc CCGGTCCTG CAGAATATTT
     251  TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
     301  GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG
     40  351  GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
     401  CCGCCTGCCT TGCGGTGCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
     451  GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG
     501  CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTTTAG
     551  CAATCGGCAT TTTTtCCCTG CAACTGAAAA AAATCATGCA AAACCGATAT
     45  601  ATCCGCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAAACT
  
```

651 TGCCGTCCTG TGGCTGTAA

This corresponds to the amino acid sequence (SEQ ID NO: 394; ORF103-1):

5  
1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPHI NRFWLILLN  
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNILYTAANL LLLFLGLYLS  
101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLA VG ILWGWLPCGL  
151 VYSASLYALG SGSAATGGLY MLAFALGTLP NLLAIGIFSL QLKIMQNR  
201 IRLCTGLSVS LWALWKLAVL WL\*

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 (SEQ ID NO: 392) shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) (SEQ ID NO: 396) from strain A of *N. meningitidis*:

15	orf103.pep	10 20 30 40 50 60	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNNTGRVSSYTAI
	orf103a	10 20 30 40 50 60	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNNTGRVSSYTAI
20	orf103.pep	70 80 90 100 110 120	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
	orf103a	70 80 90 100 110 120	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
25	orf103.pep	130 140 150 160 170 180	NPILNRLLPIKSIPACLA VG ILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP
	orf103a	130 140 150 160 170 180	NPILNRLLPIKSIPACLA VG ILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLP
30	orf103.pep	190 200 210 220	NLLAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWLX
	orf103a	190 200 210 220	NLXAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWLX

35 The complete length ORF103a nucleotide sequence (SEQ ID NO: 395) is:

1 ATGAACCANG ACATCACTTT CCTCACCCCTG TTCCTACTCG GTTCTTTCGG  
51 CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGC GC  
101 TCCAAC TCCC CCCGCATATC AACCGCTTNT GGCTGATCCT GCTGCTTAAC  
151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT  
201 CGGACAGGTC GCGGTTTCAC TCGACCAAAC CCGCGTCNTG CAGAAATATT  
251 TATACACGGC CGCCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC  
301 GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG  
351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC  
401 CCGCCTGCCT TGCCTGCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA  
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG  
501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCC AATCTTTNGG

551 CAATCGGCAT TTTTCCCTG CAACTGNAAA AAATCATGCA AAACCGATAT  
 601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAACT  
 651 TGCCGTCCTG TGGCTGTAA

5 This encodes a protein having amino acid sequence (SEQ ID NO: 396):

1 MNXDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPHI NRXLWILLN  
 51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNILYTAANL LLLFLGLYLS  
 101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLAVG ILWGWLPCGL  
 151 VYSASLYALG SGSAATGGLY MLAFALGTLP NLXAIGIFSL QLXKIMQNRY  
 201 IRLCTGLSVS LWALWKLAVL WL\*

ORF103a (SEQ ID NO: 396) and ORF103-1 (SEQ ID NO: 394) show 97.7% identity in 222 aa overlap:

15	orf103a.pep	10 20 30 40 50 60	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNTGRVSSYTAI
	orf103-1	10 20 30 40 50 60	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI
20	orf103a.pep	70 80 90 100 110 120	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
	orf103-1	70 80 90 100 110 120	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
25	orf103a.pep	130 140 150 160 170 180	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP
	orf103-1	130 140 150 160 170 180	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP
30	orf103a.pep	190 200 210 220	NLXAIGIFSLQLXKIMQNR YIRLCTGLSVSLWALWKLAVLWLX
	orf103-1	190 200 210 220	NLLAIGIFSLQLKKIMQNR YIRLCTGLSVSLWALWKLAVLWLX

#### Homology with a predicted ORF from *N.gonorrhoeae*

35 ORF103 (SEQ ID NO: 392) shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) (SEQ ID NO: 398) from *N. gonorrhoeae*:

40	orf103.pep	60	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLILLNTGRVSSYTAI
	orf103ng	60	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRISYTAI
	orf103.pep	120	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
	orf103ng	120	GLMLGLIGQLGISLDQTRVLQNILYTAASNLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
	orf103.pep	180	NPILNRLPIKSIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP



```

      |||
orf103ng      NPILNRLLPKISIPACLAVGILWGWLP CGLVYSASLYALGSGSATTGGGLYMLAFALGTLP 180
      |||

orf103.pep    NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWL 222
      |||
5  orf103ng    NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWL 222
      |||

```

The complete length ORF103ng nucleotide sequence (SEQ ID NO: 397) is:

```

1  ATGAACCACG ACATCACTTT CCTCACCTG TTCCTGCTCG GTTCTTTCGG
51 CGGAATCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGCGC
101 TCCAATCCC CCCGCATATC AACCGCTTTT GGCTGATTCT GCTGCTTAAC
151 ACAGGACGGA TAAGCAGCTA TACGGCAATC GGCCTGATGC TCGGATTAAT
201 CGGACAACTC GGCATTTCAC TCGACCAAAc ccgcgTCCTG CAAAATATTT
251 tatacacagc ctccaaCCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
301 GGTATTTCTT CCTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG
15 351 GCGCAACCTG AACCCGATAC TCAACCGGCT GCTGCCATA AAATCCATAC
401 CCGCCTGCCT TGCTGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
451 GTTTACAGCG CATCACTTTA CGCGCTGGGA AGCGGTAGTG CGACAACCGG
501 CGGACTGTAT ATGCTTGCCT TTGCACTGGG TACGCTGCC AATCTTTTGG
551 CAATCGGCAT TTTTTCCTG CAACTGAAAA AAATCATGCA AAACCGATAT
20 601 ATCCGCCTGT GTACAGGATT ATCCGTATCA TTATGGGCAT TATGGAAGCT
651 TGCCGTCCTG TGGCTGTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 398):

```

1  MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NREWLILLLN
25 51 TGRISSTAI GLMLGLIGQL GISLDQTRVL QNILYTASNL LLLFLGLYLS
101 GISSLAAKIE KIGKPIWRNL NPILNRLLPi KSIPACLAVG ILWGWLP CGL
151 VYSASLYALG SGSATTGGLY MLAFALGTL NLLAIGIFSL QLKKIMQNRY
201 IRLCTGLSVS LWALWKLAVL WL*

```

30 In addition, ORF103ng (SEQ ID NO: 398) and ORF103-1 (SEQ ID NO: 394) show 97.3% identity in 222 aa overlap:

```

      10      20      30      40      50      60
orf103-1.pep  MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhiNREWLILLLNNTGRVSSYTAI
35 orf103ng     MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhiNREWLILLLNNTGRISSTAI
      10      20      30      40      50      60

      70      80      90     100     110     120
orf103-1.pep  GLILGLIGQVGVS LDQTRVLQNILYTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
40 orf103ng     GLMLGLIGQLGISLDQTRVLQNILYTASNLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
      70      80      90     100     110     120

      130     140     150     160     170     180
orf103-1.pep  NPILNRLLPKISIPACLAVGILWGWLP CGLVYSASLYALGSGSAATGGGLYMLAFALGTLP
45 orf103ng     NPILNRLLPKISIPACLAVGILWGWLP CGLVYSASLYALGSGSATTGGGLYMLAFALGTLP
      130     140     150     160     170     180

      190     200     210     220
orf103-1.pep  NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX
      |||

```

orf103ng                      NLLAIGIFSLQLKKIMQNRYIRLCTGLSVSLWALWKLAVLWLX  
    190                      200                      210                      220

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 47

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 399):

```

10      1  ATGGAAAACC  AAAGGCCGCT  CCTAGGCTTT  CGCTTGGCAC  TTTTGGCGGC
      51  GATGACGTGG  GGAACGCTGC  CGAT.TCCGT  GCGGCAGGTA  TTGAAGTTTG
     101  TCGATGCGCC  GACGCTGGTG  TGGGTGCGTT  TTACCGTGGC  GGCGGCGGTA
     151  TTGTTTGTTT  TGCTGGCACT  GGGCGGGCGG  CTGCCGAAGC  GGCGgGGATT
     201  TTTCTTGGTG  CTCATTCAGG  CTGCTGCTGC  TCGGCGTGGC  GGGCATTTCG
     15  251  GCAAACTTTG  TGCTGATTGC  CCAAGGGCTG  CATTATATTT  CGCCGACCAC
     301  GACGCAGGTT  TTGTGGCAGA  TTTCGCCGTT  TACGATGATT  GTWGTCCGTG
     351  TGTTGGTGTT  TAAAGACCGG  ATGACTGCCG  CTCAGAAAAT  CGGCTTGGTT
     401  TTGCTGCTTG  CCGGTTTGCT  TATGTATTTT  AACGATAAAT  TCGGCGAGTT
     451  GTCGGGTTTG  GCGCGGTATG  C.AAGGGCGT  GTTGCTGTGT  GCGGCAGGCA
     20  501  GTATGGCATG  GGTGTGTAAT  GCCGTGGCGC  AAAAGCTGCT  GTCGCGCAA
     551  TTCGGGCGGC  AACAGATTCT  GCTGTTGATT  TATGCGGCAA  GTGCCGCCGT
     601  GTTCCTGCCG  TTTGCCGAAC  CGGCACACAT  CGGAAGTATG  GACGGTACGT
     651  TGGCGTGGGT  ATGTATTGCG  TATTGCTGCT  TGAATACGTT  AATCGGTTAC
     701  GGCTCGTTCG  GCGAGGCGTT  GAAACATTGG  GAGGCTTCCA  AAGTCAGCGC
     25  751  GGTAACAACC  TTGCTCCCCG  TGTTTTACCGT  AATAAATACT  TTGCTCGGGC
     801  ATTATGTGAT  GCCTGAAACT  TTTGCCGCGC  CGGA..

```

This corresponds to the amino acid sequence (SEQ ID NO: 400; ORF104):

```

30      1  MENQRPLLGF  RLALLAAMTW  GTLPXSVRQV  LKFVDAPTLV  WVRFTVAAAV
     51  LFLVLLALGGR  LPKRRDFSWC  SFRLLLLGVA  GISANFVLIA  QGLHYISPTT
     101  TQVLWQISPF  TMIVVGVLVF  KDRMTAAQKI  GLVLLLAGLL  MYFNDKFGEL
     151  SGLGAYXKGV  LLCAAGSMAW  VCNAVAQKLL  SAQFGPQQIL  LLIYAASAAV
     201  FLPFAEPAHI  GSMDGTLAWV  CIAYCCNLNTL  IGYGSFGEAL  KHWEASKVSA
     35  251  VTTLLPVFTV  INTLLGHYVM  PETFAAP...

```

Further work revealed further partial DNA sequence (SEQ ID NO: 401):

```

40      1  ATGGAAAACC  AAAGGCCGCT  CCTAGGCTTC  GCGTTGGCAC  TTTTGGCGGC
     51  GATGACGTGG  GGAACGCTGC  CGATTGCCGT  GCGGCAGGTA  TTGAAGTTTG
     101  TCGATGCGCC  GACGCTGGTG  TGGGTGCGTT  TTACCGTGGC  GGCGGCGGTA
     151  TTGTTTGTTT  TGCTGGCACT  GGGCGGGCGG  CTGCCGAAGC  GGCGGGATTTT
     201  TTCTTGGTGC  TCATTCAGGC  TGCTGCTGCT  CGGCGTGGCG  GGCATTTCGG
     251  CAAACTTTGT  GCTGATTGCC  CAAGGGCTGC  ATTATATTTT  GCCGACCACG
     301  ACGCAGGTTT  TGTGGCAGAT  TTCGCCGTTT  ACGATGATTG  TTGTCGGTGT
     351  GTTGGTGTTT  AAAGACCGGA  TGACTTGCCGC  TCAGAAAATC  GGCTTGGTTTT
     45  401  TGCTGCTTGC  CGGTTTGCTT  ATGTTTTTTA  ACGATAAATT  CGGCGAGTTG
     451  TCGGGTTTGG  GCGCGTATGC  GAAGGGCGTG  TTGCTGTGTG  CGGCAGGCAG
     501  TATGGCATGG  GTGTGTTATG  CCGTGGCGCA  AAAGCTGCTG  TCGGCGCAAT

```

551 TCGGGCCGCA ACAGATTCTG CTGTTGATTT ATGCGGCAAG TGCCGCCGTG  
 601 TTCCTGCCGT TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT  
 651 GCGTGGGTT TGTTTGCGT ATTGCTGCTT GAATACGTTA ATCGGTTACG  
 701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG  
 5 751 GTAACAACCT TGCTCCCCGT GTTACCGTA ATAwTwwCTT TGCTCGGGCA  
 801 TTATGTGATG CCTGAAACTT TTGCCGCGCC GGA...

This corresponds to the amino acid sequence (SEQ ID NO: 402; ORF104-1):

10 1 MENQRPLLGF ALALLAAMTW GTLPiAVRQV LKFVDAPTLV WVRFTVAAAV  
 51 LFVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA QGLHYISPTT  
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL  
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV  
 201 FLPPFAEPAHI GSLDGTLAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA  
 15 251 VTTLLPVFTV IXXLLGHYVM PETFAAP...

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein (SEQ ID NO: 1138) of *H. influenzae* (accession number U32769)

20 ORF104 (SEQ ID NO: 400) and HI0878 (SEQ ID NO: 1138) show 40% aa identity in 277aa overlap:

orf104 4 QRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 62  
 Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P  
 HI0878 3 QQPLLGFRTALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

25 orf104 63 --KRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 120  
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F  
 HI0878 63 LMKVRQYAW----IMLIGVIGLTSNLLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLF 118

orf104 121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180  
 K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+  
 HI0878 119 KEKLGHLQKIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178

30 orf104 181 SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL 240  
 +F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL  
 HI0878 179 LRKFNSQQILLMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCCLNTLIGYGSYAEAL 237

orf104 241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277  
 W+ SKVS V TL+P+FT++ + + HY P FAAP  
 35 HI0878 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAP 274

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF104 (SEQ ID NO: 400) shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) (SEQ ID NO: 404) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf104.pep	MENQRPLLGFRLALLAAMTWGTL	PXSVRQVLKFVDAPTLVWVRFTV	AAVLFVLLALGGR			
5	orf104a	MENQRPLLGFALALLAAMTWGTL	PIAVRQVLKFVDAPTLVWVRFTV	AAVLFVLLALGGR			
		10	20	30	40	50	60
	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFV	LIAQGLHYISPTTTQVLWQISPF	TMIVVGVLVF			
10	orf104a	LPKWRDFSWCSFRLLLLGVAGISANFV	LIAQGLHYISPTTTQVLWQISPF	TMIVVGVLVF			
		70	80	90	100	110	120
	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDK	FGELSGLGAYXKGVLLCAAGSMAW	VCNAVAQKLL			
15	orf104a	KDRMTAAQKIGLVLLLAGLLMFNDK	FGELSGLGAYAKGVLLCAAGSMAW	VCYAVAQKLL			
		130	140	150	160	170	180
	orf104.pep	SAQFGPQQIILLIYAASAAVFLPFA	EPAHIGSMDGTLAWVCIAYCCLNT	LIGYGSFGEAL			
20	orf104a	SAQFGPQQIILLIYAASAAVFLPFA	ELAHIGSLDGTALWVCFAYCCLNT	LIGYGSFGEAL			
		190	200	210	220	230	240
	orf104.pep	KHWEASKVSAVTLLPVFTVIN	TLLGHYVMPETFAAP				
25	orf104a	KHWEASKVSAVTLLPVFTVIF	SLLGHYVMPDTFAAPDMNGLGYAGAL	VVVGGAVTA	AVG		
		250	260	270	280	290	300

The complete length ORF104a nucleotide sequence (SEQ ID NO: 403) is:

	1	ATGGA	AAACC	AAAGC	CGCT	CCTAG	GCTTC	GCGTT	GGCAC	TTTGG	CGGC
30	51	GATGA	CGTGG	GGAAC	GCTGC	CGATT	GCCGT	GCGGC	CAGGTA	TGAA	AGTTTG
	101	TCGAT	GCGCC	GACGT	GCGTG	TGGGT	GCGTT	TTACC	GTTGGC	GGCGG	CGGTA
	151	TTGTT	TGTTT	TGCTG	GCATT	GGGCG	GGCGG	CTGCC	GAAGT	GGCGG	GATTT
	201	TTCTT	GGTGC	TCATT	CAGGC	TGCTG	CTGCT	CGGCG	TGGCG	GGCAT	TTCCG
	251	CAAAC	TTTGT	GCTGA	TTGCC	CAAGG	GCTGC	ATTAT	ATTTTC	GCCGA	CCACG
35	301	ACGCA	GAGTTT	TGTGG	CAGAT	TTCGC	CGTTT	ACGAT	GATTG	TTGTC	GGTGT
	351	GTTGG	TGTTT	AAAGC	CGGA	TGACT	GCCGC	TCAGA	AAATC	GGCTT	GGTTT
	401	TGCTG	CTTGC	CGGTT	TGCTT	ATGTT	TTTTTA	ACGAT	AAATT	CGGCG	AGTTG
	451	TCGGG	TTTGG	GCGCG	TATGC	GAAGG	GCGTG	TTGCT	GTGTG	CGGC	AGGCAG
	501	TATGG	CATGG	GTGTG	TATAG	CCGTG	GCGCA	AAAGC	TGCTG	TCGGC	GCAAT
40	551	TCGGG	CCGCA	ACAGAT	TCTG	CTGTT	GATTT	ATGCG	GCAAG	TGCCG	CCCGTG
	601	TTCTT	GCCGT	TTGCC	GAAC	TGAC	ATC	GGAAG	TTTGG	ACGGT	ACGTT
	651	GGCGT	GGGTT	TGTTT	TGCGT	ATTG	CTGCTT	GAAT	ACGTTA	ATCGG	TACG
	701	GCTCG	TCCGG	CGAGG	CGTTG	AAACA	TGGG	AGGCT	TCCAA	AGTCA	GCGCG
	751	GTAAC	AACCT	TGCTC	CCCGT	GTTTA	CCGTA	ATATT	TTTCTT	TGCTC	GGGCA
45	801	TTATG	TGATG	CCTGA	TA	TTGCC	CGCC	GGAT	ATGAAC	GGTTT	GGGTT
	851	ATGCC	GCGCG	ACTGG	TCGTG	GTCGG	GGGTG	CGGTT	ACGGC	GGCGG	TGGG
	901	GACAG	GCTGT	TCAAC	GCCG	CTAG					

This encodes a protein having amino acid sequence (SEQ ID NO: 404):

	1	MENQRPLLGF	ALALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
50	51	LFVLLALGGR	LPKWRDFSWC	SFRLLLLGVA	GISANFVLIA	QGLHYISPTT
	101	TQVLWQISPF	TMIVVGVLVF	KDRMTAAQKI	GLVLLLAGLL	MFNDKFGEL
	151	SGLGAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQQIL	LLIYAASAAV

201 FLPFAELAHIGSLDGTALWVCFAYCCLNTLIGYGSFGEALKHWEASKVSA  
 251 VTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG  
 301 DRLFKR\*

5 ORF104a (SEQ ID NO: 404) and ORF104-1 (SEQ ID NO: 402) show 98.2% identity in 277 aa overlap:

		10	20	30	40	50	60
	orf104a.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
10	orf104-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR					
		10	20	30	40	50	60
	orf104a.pep	LPKWRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF					
15	orf104-1	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF					
		70	80	90	100	110	120
	orf104a.pep	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
20	orf104-1	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL					
		130	140	150	160	170	180
	orf104a.pep	SAQFGPQQILLIYAASAAVFLPFAELAHIGSLDGTALWVCFAYCCLNTLIGYGSFGEAL					
25	orf104-1	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGTALWVCFAYCCLNTLIGYGSFGEAL					
		190	200	210	220	230	240
	orf104a.pep	KHWEASKVSAVTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG					
30	orf104-1	KHWEASKVSAVTLLPVFTVIXLLGHYVMPETFAAP					
		250	260	270			

#### Homology with a predicted ORF from *N. gonorrhoeae*

ORF104 (SEQ ID NO: 400) shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) (SEQ ID NO: 406) from *N. gonorrhoeae*:

35	orf104.pep	MENQRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orf104ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF	120
40	orf104ng	LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF	120
	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL	180
	orf104ng	KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180

```

orf104.pep  SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL  240
             |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf104ng    SAQFGPQQIILLIYAASAAVFLXAEPAHIGSLDGTALAWVCFVYCCLNTLIGYGSFGEAL  240

orf104.pep  KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP  277
             |||||:|||||:|||||:|||||
5 orf104ng   KHWEASKVSAVTTLLPVFTVIFSLLLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG  300

```

The complete length ORF104ng nucleotide sequence (SEQ ID NO: 405) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 406):

```

10      1  MENQRPLLGF ALALLAAMTW  GTLPIAVRQV  LKFVDAPTLV  WVRFTVAAAV
      51  LFVLLALGGR LPKRRDFSWH  SFRLLLLGVT  GISANFVLIA  QGLHYISPTT
     101  TQVLWQISPF TMIVVGVLVF  KDRMTAAQKI  GLVLLLVGLL  MFFNDKFGEL
     151  SGLGAYAKGV LLCAAGSMAW  VCYAVAQKLL  SAQFGPQQIL  LLIYAASAAV
     201  FLLXAEPAHI GSLDGTALAV  CFVYCCLNTL  IGYGSFGEAL  KHWEASKVSA
15      251  VTTLLPVFTV IFSLLLGHYV  PDTFAAPDMN  GLGYVGALVV  VGGAVTAAVG
     301  DRPFKRR*

```

Further work revealed the complete gonococcal nucleotide sequence (SEQ ID NO: 407):

```

20      1  ATGGAAAACC  AAAGCCGCT  CCTAGGCTTC  GCGTTGGCAC  TTTTGGCGGC
      51  GATGACGTGG  GGGACGCTGC  CGATTGCCGT  GCGGCAGGTA  TTGAAGTTTG
     101  TCGATGCGCC  GACGCTGGTG  TGGGTGCGTT  TTACCGTGGC  GGCGGCGGTA
     151  TTGTTTGTTT  TGCTGGCATT  GGGCGGGCGG  CTGCCGAAGC  GGCGGGATTT
     201  TTCTTGGCAT  TCATTCAGGC  TGCTGCTGCT  CGCGCTGACG  GGCATTTCGG
     251  CAAACTTTGT  GCTGATTGCC  CAAGGGCTGC  ATTATATTTT  GCCGACCACG
25      301  ACGCAGGTTT  TGTGGCAGAT  TTCGCCGTTT  ACGATGATTG  TTGTCGGCGT
     351  GTTGGTGTTT  AAAGACCGGA  tgaCTGCCGC  GCAGAAAATC  GGTTTGGTTT
     401  TGCTGCTtgT  CGGTttgCTT  ATGTTTtTa  ACGACAAATT  CGGCGAGTTG
     451  TCGGGTTTGG  GCGCGTATGC  GAAGGGCGTG  TTGCTGTGTG  CGGCAGGCAG
     501  TATGGCCTGG  GTGTGTTATG  CCGTGGCGCA  AAAGCTGCTG  TCGGCGCAAT
30      551  TCGGGCCGCA  ACAGATTCTG  CTGTTGATTT  ATGCGGcaag  tgccgccGTG
     601  TTCctgcccT  TTGccgaaCC  GGCACACATC  GGAAGTTTgg  aCGGTACGtt
     651  GGCCTGGGTT  TGTTTGTGT  ATTGCTGCTT  GAATACGTTA  ATCGGTACG
     701  GCTCGTTCGG  CGAGGCGTTG  AAACATTGGG  AGGCTTCCAA  AGTCAGCGCG
     751  GTAACAACCT  TGCTCCCCGT  GTTTACCGTA  ATATTTTCTT  TGCTCGGGCA
35      801  TTATGTGATG  CCTGATACTT  TTGCCGCGCC  GGATATGAAC  GGTTTTGGGTT
     851  ATGTCGGCGC  ACTGGTCGTG  GTCGGGGGTG  CGGTTACGGC  GGCGTTGGGG
     901  GACAGGCCGT  TCAAACGCCG  CTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 408; ORF104ng-1):

```

40      1  MENQRPLLGF ALALLAAMTW  GTLPIAVRQV  LKFVDAPTLV  WVRFTVAAAV
      51  LFVLLALGGR LPKRRDFSWH  SFRLLLLGVT  GISANFVLIA  QGLHYISPTT
     101  TQVLWQISPF TMIVVGVLVF  KDRMTAAQKI  GLVLLLVGLL  MFFNDKFGEL
     151  SGLGAYAKGV LLCAAGSMAW  VCYAVAQKLL  SAQFGPQQIL  LLIYAASAAV
     201  FLPFAEPAHI GSLDGTALAV  CFVYCCLNTL  IGYGSFGEAL  KHWEASKVSA
45      251  VTTLLPVFTV IFSLLLGHYV  PDTFAAPDMN  GLGYVGALVV  VGGAVTAAVG
     301  DRPFKRR*

```

ORF104ng-1 (SEQ ID NO: 408) and ORF104-1 (SEQ ID NO: 402) show 97.5% identity in 277 aa overlap:

-321-

5	orf104-1.pep	10 20 30 40 50 60	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
	orf104ng-1	10 20 30 40 50 60	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
10	orf104-1.pep	70 80 90 100 110 120	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF
	orf104ng-1	70 80 90 100 110 120	LPKRRDFSWHSFRLLLLLGVGTGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF
15	orf104-1.pep	130 140 150 160 170 180	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
	orf104ng-1	130 140 150 160 170 180	KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
20	orf104-1.pep	190 200 210 220 230 240	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
	orf104ng-1	190 200 210 220 230 240	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL
25	orf104-1.pep	250 260 270	KHWEASKVSAVTTLPLPVFTVIXLLGHYVMPETFAAP
	orf104ng-1	250 260 270 280 290 300	KHWEASKVSAVTTLPLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGAVTAAVG

In addition, ORF104ng-1 (SEQ ID NO: 408) shows significant homology with a hypothetical *H.influenzae* protein (SEQ ID NO: 1138):

30	gi 1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306		
	Score = 237 bits (598), Expect = 8e-62		
	Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)		
	Query: 30	QRPXXXXXXXXXXMTWGTLP	IAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXXP- 88
		Q+P	M WG+LPIA++QVL ++A T+VW P
	Sbjct: 3	QQPLLGF	TFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSL LALLAYKKQLPE 62
35	Query: 89	--KRRDFSWHSFRLLLL	LGVTGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF 146
		K R ++W ++L+GV	G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
	Sbjct: 63	LMKVRQYAW---	IMLIGVIGLT SNFLFSSSLN YIEPSVAQIFIHLSSFGMLICGVLIF 118
40	Query: 147	KDRMTAAQKIXXXXXXXXXXX	MFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206
		K+++ QKI	+FFND+F +GL Y+ GV+L G++ WV Y +AQKL+
	Sbjct: 119	KEKLGLHQKIGLFL	LLIGLGLFFNDRF DAFAGLNQYSTGVILGVG GALI WVAYGMAQKLM 178
	Query: 207	SAQFGPQQIILLIYAASAAV	FLPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL 266
		+F QQILL++Y A F+P A+ + +	L LA +CF+YCCCLNTLIGYGS+ EAL
	Sbjct: 179	LRKFNSQQILLMMYL	GCAIAFMPMADFSQVQELT-PLALICFIYCCCLNTLIGYGSYAEAL 237
45	Query: 267	KHWEASKVSAVTTLPLPV	FTVIFSLLGHYVMPDTFAAPDMN 306
		W+ SKVS V TL+P+FT++FS	+ HY P FAAP++N
	Sbjct: 238	NRWDVSKVSVVITLV	PLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 48

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 409):

```

1  ATGGTAGCTC GTCGGGCTCA TAACCCGAAG GTCGTAGGTT CGAATCCTGT
51  .CCC GCAACC TAATTTCAAA CCCCTCGGTT CAATGCCGAG GG.GTTTTGT
101 T.TTGCTGT TTCCTGTTTC CTGTTTCCTG CCGCCTCCGT TTTTGCCGG
151 ATTTTCCTTC CGGCCGAAT ATCGGAACGG CAGACCGCCG TCTGTTGCG
201 GTTGCAAATT CAGGCAGTTT GGCTACAATC TTCCGCATTG TCTTCAAGAA
251 AGCCAACCAT GCCGACCGTC CGTTTACC G AATCCGTCAG CAAACAAGAC
301 CTTGATGCTC TGTTGAGTG GGC AAAAGCA AGTTACGGTG CAGAAAGTTG
351 CTGGA A ACG CTGTATCTGA ACGGTCysCC TTTGGGCAAC CTGTCGCCGG
15 401 AATGGGTGGA ACGCGTsmmA AAAGACTGGG AGGCAGGCTG CyCGGAGTCT
451 TCAGACGGCA TTTTCTGAA TgCGGACGGc TGgCctGATA TGGgCGGAcg
501 cTTACAGCAC CTCGCCCTCG GTTGGCACTG TGCGGGGCTG TTGGAGCGsT
551 GGCGCAACGA GTGTTTCGAC CTGACCGACG GCGGCGGCAA CCCCTGTGTTT
601 ACGCTCGaAc GCGCCGyTTT mCGTCCTkTC GGACTGCTCA GCCGCGCCGT
20 651 CCATCTCAAC GGTCTGACCG AATCGGACGG CCGATGGCAT TTCTGGATAG
701 GCAGGCGCAG TCCGCACAAA GCAGTCGATC CCAACAAACT CGACAATACT
751 rCCGCCGGCG GTGTTTCCGG CGGCGAAATG CCGTCTGAAG CCGTGTGTCTG
801 CGAAAGCAGC GAAGAAGCCG GTTTGGATAA AACGCTGcTT CCGCTCATCC
851 GCCCGGTATC GCAGCTGCAC AGCCTGCGCT CCGTCAGCCG GGTGTACAC
25 901 AATGAAATCC TGTATGTATT CGATGCCGTC CTGCCG...
```

This corresponds to the amino acid sequence (SEQ ID NO: 410; ORF105):

```

1  MVARRAHNPV VGSNPXPAT XFQTPRFNAE XVLXLPVSCF LFPAASVFRC
51  IFLPAAISER QTAVCLRLQI QAVWLQSSAL SSRKPTMPTV RFTESVSKQD
30 101 LDALFEWAKA SYGAESCWKT LYLNGLPLGN LSPEWVERVX KDWEAGCXES
151 SDGIFLNADG WPDMMGRLQH LALGWHCAGL LDGWRNECFD LTDGGGNPLF
201 TLERAXRPX GLLSRAVHLN GLTESDGRWH FWIGRRSPHK AVDPNKLDNT
251 XAGGVSGGEM PSEAVCRESS EEAGLDKTL PLIRPVSQLH SLRSVSRGVH
35 301 NEILYVFDAV LP...
```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 411):

```

1  ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
51  TCTGTTGAG TGGGCAAAAG CAAGTTACGG TGCAGAAAGT TGCTGGAAAA
40 101 CGCTGTATCT GAACGGTCTG CCTTTGGGCA ACCTGTCGCC GGAATGGGTG
151 GAACGCGTCA AAAAAGACTG GGAGGACGGC TGCTCGGAGT CTTACAGCGG
201 CATTTTCTG AATGCGGACG GCTGGCCTGA TATGGGCGGA CGCTTACAGC
251 ACCTCGCCCT CGGTGGCAC TGTGCGGGC TGTGGACGG CTGGCGCAAC
301 GAGTGTTTCG ACCTGACCGA CGGCGGCGGC AACCCTTGT TCACGCTCGA
351 ACGCGCCGCT TTCCGTCCTT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
45 401 ACGGTCTGAC CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGGCGC
451 AGTCCGCACA AAGCAGTCGA TCCCAACAAA CTCGACAATA CTGCCGCCGG
501 CGGTGTTTCC GCGGCGGAAA TGCCGTCTGA AGCCGTGTGT CGCGAAAGCA
```



551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCC GGTA  
 601 TCGCAGCTGC ACAGCCTGCG CTCCGTCAGC CGGGGTGTAC ACAATGAAAT  
 651 CCTGTATGTA TTCGATGCGG TCCTGCCCGA AACCTTCCTG CCTGAAAATC  
 701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CCGTCTGTTG  
 751 GATGCCATGT TGTCGGGAAA CATGATGCAC GACGCGCAAC TGGTTACGCT  
 801 GGACGCGTTT TGCCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG  
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence (SEQ ID NO: 412; ORF105-1):

1 MPTVRFTESEV SKQDL DALFE WAKASYGAES CWKTLYL NGL PLGNLSPEWV  
 51 ERVKKDWEAG CSESSDGIFL NADGW PDMGG RLQHLALGWH CAGLLDGWRN  
 101 ECFDLTDGGG NPLFTLERAA FRPFG LLSRA VHLNGLTESD GRWHFWIGRR  
 151 SPHKAVDPNK LDNTAAGGVS GGEMPSEAVC RESSEEAGLD KTLPLIRPV  
 201 SQLHSLRSVS RGVHNEILYV FDAVLPE TFL PENQDGEVAG FEKMDIGGLL  
 251 DAMLSGNMMH DAQLVTLD AF CRYGLIDAAH PLSEWLDGIR L\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 (SEQ ID NO: 410) shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) (SEQ ID NO: 414) from strain A of *N. meningitidis*:

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESEVSKQDL DALFEWAKASYGAES					
orf105a				MPTVRFTESEVSKHDL DALFEWAKASYGAES		
				10	20	30
	120	130	140	150	160	170
orf105.pep	CWKTLYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGIFLNADGW PDMGGRLQHLALGWH					
orf105a	CWKTLYL NGLPLGNLSPEWAERVKKDWEAGCSESSDGIFLNADGW PDMGGRRQLHLARIWK					
	40	50	60	70	80	90
	180	190	200	210	220	230
orf105.pep	CAGLLDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRR					
orf105a	EAGLLHGWRDECFDLTDGGSNPLFALERAAFRPFG LLSRAVHLNGLVESDGRWHFWIGRR					
	100	110	120	130	140	150
	240	250	260	270	280	290
orf105.pep	SPHKAVDPNKLDNTXAGGVSSGGEMPSEAVCRESEEAGLDKTLPLIRPVSQLHSLRSVS					
orf105a	SPHKAVDPKLDNTAAGGVSSGELPSETVCRESEEAGLDKTLPLIRPVSQLHSLRPVS					
	160	170	180	190	200	210
	300	310				
orf105.pep	RGVHNEILYV FDAVL P					
orf105a	RGVHNEILYV FDAVLPE TFLPENQDGEVAGFEKMDIGGLLAAMLSGNMMHDAQLVTLD AF					
	220	230	240	250	260	270

The complete length ORF105a nucleotide sequence (SEQ ID NO: 413) is:

```
1  ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAAACACG ACCTTGATGC
51 CCTATTCGAG TGGGCAAAGG CAAGTTACGG TCGGGAAGT TGCTGGAAAA
101 CGCTGTATCT GAACGGTCTG CTTTGGGCA ATCTGTCGCC GGAATGGGCG
5   151 GAGCGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTAGACGG
201 CATTTTCCTG AATGCGGACG GCTGGCCAGA TATGGGCAGA CGCTTGCAGC
251 ACCTCGCCCG AATATGGAAA GAAGCGGGAC TGCTTACGG CTGGCGCGAC
301 GAGTGTTTCG ACCTGACCGA CGGCGGCAGC AATCCCTTGT TCGCGCTCGA
351 ACGCGCCGCT TTCCGTCCGT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
10  401 ACGGTTTGGT CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGGCGC
451 AGTCCGCACA AAGCAGTCGA TCCCACAAA CTCGACAATA CTGCCGCCGG
501 CGGTGTTTCC AGCGGTGAAT TGCCGTCTGA AACCGTGTGT CGCGAAAGCA
551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGGTA
601 TCGCAGCTGC ACAGCCTGCG CCCCCTCAGC CGGGGTGTGC ACAATGAAAT
15  651 CCTGTATGTA TTCGATGCCG TCCTGCCCGA AACCTTCTG CCTGAAAATC
701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTTG
751 GCTGCCATGT TGTGGGAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
801 GGACGCGTTT TGCCGTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
20  851 AGTGGCTGGA CGGCATACGT TTATAG
```

This encodes a protein having amino acid sequence (SEQ ID NO: 414):

```
1  MPTVRFTESV SKHDLDALFE WAKASYGAES CWKTLYLNL PLGNLSPWEA
51  ERVKKDWEAG CSESSDGIFL NADGWPMGR RLQHLARIWK EAGLLHGWRD
25  101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV
201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGLL
251 AAMLSGNMMH DAQLVTLD AF CRYGLIDAAH PLSEWLDGIR L*
```

ORF105a (SEQ ID NO: 414) and ORF105-1 (SEQ ID NO: 412) show 93.8% identity in 291 aa overlap:

```
10      20      30      40      50      60
orf105a.pep MPTVRFTESVSKHDLDALFEWAKASYGAESCWKTLYLNLPLGNLSPWEAERVKKDWEAG
35  orf105-1  MPTVRFTESVSKQDLDALFEWAKASYGAESCWKTLYLNLPLGNLSPWVERVKDWEAG
      10      20      30      40      50      60
70      80      90      100     110     120
orf105a.pep CSESSDGIFLNADGWPMGRRLQHLARIWKEAGLLHGWRDECFDLTDGGSNPLFALERAA
40  orf105-1  CSESSDGIFLNADGWPMGGRRLQHLALGWHCAGLLDGWRNECFDLTDGGGNPLFLTERAA
      70      80      90      100     110     120
130     140     150     160     170     180
orf105a.pep FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGELPSETVC
45  orf105-1  FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLNTAAGGVSGGEMPSEAVC
      130     140     150     160     170     180
190     200     210     220     230     240
orf105a.pep RESSEEAGLDKTLPLIRPVSQHLHSLRPVSRGVHNEILYVDAVLPETFLPENQDGEVAG
50  orf105-1  RESSEEAGLDKTLPLIRPVSQHLHSLRSVSRGVHNEILYVDAVLPETFLPENQDGEVAG
      190     200     210     220     230     240
```

-325-

		250	260	270	280	290
orf105a.pep		FEKMDIGGLLAAMLSGNMMHDAQLVTLD	AFCRYGLIDAAHPLSEWLDGIRLX			
orf105-1		FEKMDIGGLLDAMLSGNMMHDAQLVTLD	AFCRYGLIDAAHPLSEWLDGIRLX			
5		250	260	270	280	290

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 (SEQ ID NO: 410) shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) (SEQ ID NO: 416) from *N. gonorrhoeae*:

10	orf105.pep	MVARRAHNPKVVGSNPXPATXFQTPRFNAEXVLKLPVSCFLFPAASVFCRIFLPAAISER	60
	orf105ng	MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER	55
	orf105.pep	QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	120
	orf105ng	QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWKT	115
15	orf105.pep	LYLNGXPLGNLSPEWVERVXKDWEAGCXESSDGI FLNADGWPDMMGRLQHLALGWHCAGL	180
	orf105ng	LYLNRLPLGNLSPEWAERIKKDWEAGCSESSNGI FLNADGWPDMMGRLQHLARTWNKAGL	175
	orf105.pep	LDGWRNECFDLTDGGGNPLFTLERAXRXPXGLLSRAVHLNGLTESDGRWHFWIGRRSPHK	240
20	orf105ng	LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK	235
	orf105.pep	AVDPNKLDNTXAGGVSGGEMPSEAVCRESSEEEAGLDKTLPLIRPVSQHLSLRVSRRGVH	300
	orf105ng	AVDPGKLDNIAGGVSGGEMPSEAVCRESSEEEAGLDKTLFPLIRPVSRLHSLRVPVSRGVH	295
25	orf105.pep	NEILYVFDAVLP	312
	orf105ng	NEILYVFDAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVTLDAPYRYG	355

A complete length ORF105ng nucleotide sequence (SEQ ID NO: 415) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 416):

30	1	MVARRAHNPK	VVGSNPAPAT	KYQTPRFNAE	GVLFFLFPAA	SVFCRIFLPA
	51	AI SERQAAVC	LRLQIQAVWL	QSSALCSRKP	AMPTVRFTES	VSKQDLDALE
	101	ERAKASYGAE	SCWKTLYLNR	LPLGNLSPEW	AERIKKDEWA	GCSESSNGIF
	151	LNADGWPDMD	GRLQHLARTW	NKAGLLHGWR	NECFDLTDGG	GNPLFTLERA
	201	AFRPFGLLIR	AVHLNGLVES	NGRWHFWIGR	RSPHKAVDPG	KLDNIAGGGV
35	251	SGGEMPSEAV	CRESSEEEAGL	DKTLFPLIRP	VSRLHSLRPV	SRGVHNEILY
	301	VFDAVLPETF	LPENQDGEVA	GFEKMDIGGL	LDAMLSKNMM	HDAQLVTLDA
	351	FYRYGLIDAA	HPLSEWLDGI	RL*		

Further work revealed the complete nucleotide sequence (SEQ ID NO: 417):

40	1	ATGCCGACCG	TCCGTTTTAC	CGAATCCGTC	AGCAAACAAG	ACCTTGATGC
	51	CCTGTTTCGAG	CGGGCAAAAG	CAAGTTACGG	TGCCGAAAGT	TGCTGGAAAA
	101	CGCTGTATCT	GAACCGTCTT	CCTTTGGGCA	ATCTGTCCGC	GGAATGGGCT
	151	GAGCGCATCA	AAAAAGACTG	GGAGGCAGGC	TGCTCCGAGT	CTTCAGACGG

201 CATTCTTCTG AATGCGGACG GCTGGCCGGA TATGGGCGGA CGCTTGCAGC  
251 ACCTCGCCCG CACATGGAAC AAGGCGGGGC TGCTTCACGG ATGGCGCAAC  
301 GAGTGTTCG AÇCTGACCGA CGGCGGCGGC AACCCCTTGT TCACGCTCGA  
351 ACGCGCCGCT TTCCGTCCGT TCGGACTACT CAGCCGCGCC GTCCATCTCA  
401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC  
451 AGTCCGCACA AAGCAGTCGa tCCCGCAAG CTCGACAATA TTGCCGCGCG  
501 CCGTGTTCCT GCGCGCGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA  
551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA  
601 TCGCGGCTGC ACAGCCTTCG CCCCCTCAGC CGAGGTGTGC ACAATGAAAT  
651 CCTGTATGTG TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAAATC  
701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG  
751 GATGCCATGT TGTGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT  
801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG  
851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence (SEQ ID NO: 418; ORF105ng-1):

1 MPTVRFTESV SKQDLDFE RAKASYGAES CWKTLNRL PLGNLSPEWA  
51 ERIKDWEG CSESSDGIFL NADGWPMGG RLQHLARTWN KAGLLHGWRN  
101 ECFDLTDGGG NPLFTLRAA FRPFGLLSRA VHLNGLVESN GRWHFWIGRR  
151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV  
201 SRLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGLL  
251 DAMLSKNMMH DAQLVTLDFA YRYGLIDAAH PLSEWLDGIR L\*

ORG105ng-1 (SEQ ID NO: 418) and ORF105-1 (SEQ ID NO: 412) show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105-1.pep	MPTVRFTESVSKQDLDFE	WAKASYGAESC	WKTLNRLPLGNLSPEW	VERVKDWEAG			
orf105ng-1	MPTVRFTESVSKQDLDFE	RAKASYGAESC	WKTLNRLPLGNLSPEWA	ERIKDWEG			
	10	20	30	40	50	60	
		70	80	90	100	110	120
orf105-1.pep	CSESSDGIFLNADGWPMGG	RLQHLALGWHCAGLLD	GWNECFDLTDGGGNPL	FTLRAA			
orf105ng-1	CSESSDGIFLNADGWPMGG	RLQHLARTWNKAGLLH	GWNECFDLTDGGGNPL	FTLRAA			
	70	80	90	100	110	120	
		130	140	150	160	170	180
orf105-1.pep	FRPFGLLSRAVHLNGLT	ESDGRWHFWIGRRSPH	KAVDPNKLNTAAGGV	SGGEMPSEAVC			
orf105ng-1	FRPFGLLSRAVHLNGL	VESNGRWHFWIGRRSPH	KAVDPKLDNIAGGGV	SGGEMPSEAVC			
	130	140	150	160	170	180	
		190	200	210	220	230	240
orf105-1.pep	RESSEEAGLDKTLPLIR	PVSQHLHSLRSVSRG	VHNEILYVFDAVLPET	FLPENQDGEVAG			
orf105ng-1	RESSEEAGLDKTLPLIR	PVSRLHSLRPVSRG	VHNEILYVFDAVLPET	FLPENQDGEVAG			
	190	200	210	220	230	240	
		250	260	270	280	290	
orf105-1.pep	FEKMDIGLLDAMLSGN	MMHDAQLVTLDFA	CRYGLIDAAHPLSEW	LDGIRLX			
orf105ng-1	FEKMDIGLLDAMLSKN	MMHDAQLVTLDFA	RYGLIDAAHPLSEW	LDGIRLX			
	250	260	270	280	290		

Furthermore, ORF105ng-1 (SEQ ID NO: 418) shows homology with a yeast enzyme(SEQ ID NO: 1139):

```

5      sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
      )gi|1076928|pir||S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
      (Schizosaccharomyces pombe) )gi|666111 (X84417) thiamin pyrophosphokinase
      [Schizosaccharomyces pombe] )gi|2330852|gnl|PID|e334056 (Z98533) thiamin
      pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
      Score = 105 bits (259), Expect = 4e-22
      Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

10     Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERA AFRPFGLLSRAVHLNGLVESNGRW--HFWI 441
           N  G+  WRNE + +      P+  +ER  F  FG LS  VH      + +      W+
      Sbjct: 96  NTFGIADQWRNELYTVYGKSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

      Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRLHSLR 621
           RRSP K  P  LDN  GG++ G+      + +E SEEA LD +  LI P  +  ++
15     Sbjct: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNL-PCGTVSYIK 214

      Query: 622 PVSRG-VHNEILYVFDVAVLPETFLPENQDGEVAGFEKMDIGLLDAML SKNMHDAQLVT 798
           R  +  E+ YVFD  + +  +P  DGEVAGF  + +  +L  +  K+  +  LV
      Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPRINDGEVAGFSLPLNQLHELELKSFKPNCALVL 274

      Query: 799 LDAFYRYGLIDAAHP 843
           LD  R+G+I  HP
20     Sbjct: 275 LDFLIRHGIITPQHP 289

```

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 419):

```

30      1  ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
      51  CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
      101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
      151 TTGATATTTG GTAACATATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
      201 ACCTGCATCG GCGTAATCA GGGTGATATGC ACCGgATACG rGkACAATTA
      251 CAGCGAAATT CGTGGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
35     301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
      351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAAC TGG
      401 GTCGTCTGAA GCTGATACAC GGGAATGAAA CGCGCAGCcT TAAAGCAACT
      451 GTCGAACGTT TGGAAAACCA GGAAC TCCAT ATTTTCGCAAC AGATAGACGG
      501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
40     551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 420; ORF107):

```

1  MNRPKQPPFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51 LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTLAE QELGRLKLIH GNETRSLKAT
151 VERLENQELH ISQQIDGQKR RIRLAEEMLQ KYRFLSXQ*

```

5

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF107 (SEQ ID NO: 420) shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) (SEQ ID NO: 422) from strain A of *N. meningitidis*:

```

10      10      20      30      40      50      60
orf107.pep MNRPKQPPFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT
orf107a    MNRPKQPPFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT
          10      20      30      40      50      60

15      70      80      90      100     110     120
orf107.pep TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT
orf107a    TVEGQILPASGVIRVYAPDTGTITAKFXEDGEKVKAGDKLFALSTSRFGAGDSVQQQLKT
          70      80      90      100     110     120

20      130     140     150     160     170     180
orf107.pep EAVLKKTLAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ
orf107a    EAVLKKTLAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ
          130     140     150     160     170     180

25      189
orf107.pep KYRFLSXQX
orf107a    KYRFLSANDAVPKQEMMNVKAELLEQAKLDAYRREEVGLLQEIRTQNLTLSLPAAX
          190     200     210     220     230

30

```

The complete length ORF107a nucleotide sequence (SEQ ID NO: 421) is:

```

1  ATGAATAGAC CCAAGCAACC NTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51 CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
35 151 TTGATATTTG GTAACATATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
201 ACCTGCATCG GCGGTAATCA GGGTGTATGC ACCGATACG GGGACAATTA
251 CNGCGAAATT CNTGGAAGAT GGAGAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGATAGCG TGCAGCAGCA
351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAACTGG
40 401 GTCGTCTGAA GCTGATACAC GGGAATGAAA CGCGCAGCCT TAAAGCAACT
451 GTCGAACGTT TGGAAAACCA GGAATCCAT ATTTGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC CAATGATGCA GTGCCAAAAC AAGAAATGAT GAATGTCAAG
601 GCAGAGCTTT TAGAGCAGAA AGCCAACTT GATGCCTACC GCCGAGAAGA
45 651 AGTCGGGCTG CTTCAGGAAA TCCGCACGCA GAATCTGACA TTGGNNAGCC
701 TCCCCCAAGC GGCATGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 422):

5           1   MNRPKQPFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF  
          51   LIFGNYTRKT TVEGQILPAS GVIRVYAPDT GTITAKFXED GEKVKAGDKL  
         101   FALSTSRFGA GDSVQQQLKT EAVLKKTAE QELGRLKLIH GNETRSLKAT  
         151   VERLENQELH ISQQIDGQKR RIRLAEEMLQ KYRFLSANDA VPKQEMMNVK  
         201   AELLEQKAKL DAYRREEVGL LQEIRTQNL LXSLPQAA\*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 (SEQ ID NO: 420) shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) (SEQ ID NO: 424) from *N. gonorrhoeae*:

10           orf107.pep   MNRPKQPFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT   60  
                          |||||:|||||  
          orf107ng   MNRPKQPFFRPEVAIARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT   60  
  
          orf107.pep   TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT   120  
                          |:|||||  
15           orf107ng   TMEGQILPASGVIRVYAPDTGTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT   120  
  
          orf107.pep   EAVLKKTAEQELGRLKLIHGNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLQ   180  
                          |||||:|||||  
          orf107ng   EAVLKKTAEQELGRLKLIHENETRSLKATVERLENQKLHISQQIDGQKRRIRLAEEMLR   180  
  
          orf107.pep   KYRFLSXQ   188  
                          |||||  
20           orf107ng   KYRFLSAQ   188

The complete length ORF107ng nucleotide sequence (SEQ ID NO: 423) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 424):

25           1   MNRPKQPFFR PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF  
          51   LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL  
         101   FALSTSRFGA GGSVQQQLKT EAVLKKTAE QELGRLKLIH ENETRSLKAT  
         151   VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ\*

30   Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 50**

35   The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 425):

1   ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC

5

10

51	GTGCGGCAAA	TCCGTAATA	CGGCGGTACA	GCCGCAAAAC	GCGGTACAAA
101	GCGCGCCGAA	ACCGGTTTTC	AAAGTCATAT	ATATCGACAA	TACGGCGATT
151	GCCGGTTTGG	ATTTGGGACA	AAGCAGCGAA	GGCAAAACCA	ACGACGGCAA
201	AAAACAAATC	AGTTATCCGA	TTAAAGGCTT	GCCGGAACAA	AATGTTATCC
251	GACTGATCGG	CAAGCATCCC	GGCGACTTGG	AAGCCGTCAG	CGGCAAATGT
301	ATGGAAACCG	ATGATAAGGA	CAGTCCGGCA	GGTTGGGCAG	AAAACGGCGT
351	GTGCCATACC	TTGTTTGCCA	AACTGGTGCG	CAATATCGCG	GAAGACGGCG
401	GCAAACGTAG	GGATTACCTA	GTTTCGCATG	CCGCCCTGCA	ACCCATCCAG
451	GCAAGGCAAA	GCGGCTATGC	CGCCGTGCAG	AACGGACGCT	ATGTGCTGGA
501	AATCGACAGC	GAAGGGGCGT	TTTATTTCCT	CCGCCGCCAT	TATTGA

This corresponds to the amino acid sequence (SEQ ID NO: 426; ORF108):

15

1	MLNTFFAVLG	GCLLXLPCGK	SVNTAVQPQN	AVQSAPKPVF	KVIYIDNTAI
51	AGLDLQGSSE	GKTNDGKKQI	SYPIKGLPEQ	NVIRLIGKHP	GDLEAVSGKC
101	METDDKDSPA	GWAENGVCHT	LFAKLVGNIA	EDGGKLT DYL	VSHAALQPYQ
151	AGKSGYAAVO	NGRYVLEIDS	EGAFYFRRRH	Y*	

Further work revealed the following DNA sequence (SEQ ID NO: 427):

20	1	ATGCTGAAAA	CATCTTTTGC	CGTATTGGGC	GGCTGCCTGC	TGCTTGCCGC
	51	CTGCGGCAAA	TCCGAAAATA	CGGCGGAACA	GCCGCAAAAC	GCGGTACAAA
	101	GCGCGCCGAA	ACCGGTTTTC	AAAGTCAAAT	ATATCGACAA	TACGGCGATT
	151	GCCGGTTTGG	ATTTGGGACA	AAGCAGCGAA	GGCAAAACCA	ACGACGGCAA
	201	AAAACAAATC	AGTTATCCGA	TTAAAGGCTT	GCCGGAACAA	AATGTTATCC
25	251	GAAGTATCGG	CAAGCATCCC	GGCGACTTGG	AAGCCGTFCAG	CGGCAAAATGT
	301	ATGGAACCG	ATGATAAGGA	CAGTCCGGCA	GGTTGGGCAG	AAAACGGCGT
	351	GTGCCATACC	TTGTTTGCCA	AACTGGTGGG	CAATATCGCC	GAAGACGGCG
	401	GCAAACCTGAC	GGATTACCTA	GTTTCGCGATG	CCGCCCTGCA	ACCTATATCAG
	451	GCAGGCAAAA	GCGGCTATGC	CGCCGTGCAG	AACGGACGCT	ATGTGCTGGA
30	501	AATCGACAGC	GAAGGGGCGT	TTTATTTCCT	CCGCCGCCAT	TATTGA

This corresponds to the amino acid sequence (SEQ ID NO: 428; ORF108-1):

35

1	MLKTSFAVLG	GCLLLAACGK	SENTAEQPQN	AVQSAPKPVF	KVKYIDNTAI
51	AGLDLGQSSE	GKTNDGKKQI	SYPIKGLPEQ	NVIRLIGKHP	GDLEAVSGKC
101	METDDKDSPA	GWAENGVCHT	LFAKLVGNIA	EDGGKLT DYL	VSHAALQPYQ
151	AGKSGYAAVQ	NGRYVLEIDS	EGAFYFRRRH	Y*	

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 (SEQ ID NO: 428) shows 88.4% identity over a 181aa overlap with a predicted ORF  
40 (ORF108.ng) (SEQ ID NO: 430) from *N. gonorrhoeae*:

	orf108.pep	MLNTFFAVLGGCLLXLPCGKSVENTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLQGSSE	60
	orf108ng	MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALQGSSE	60
45	orf108.pep	GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSAPAGWAENGVCHT	120
	orf108ng	GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT	120



```

orf108.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH 181
|||||
orf108ng    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH 181

```

ORF108-1 (SEQ ID NO: 428) shows 92.3% identity with ORF108ng (SEQ ID NO: 430) over the same 181 aa overlap:

```

orf108-1.pep  MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE 60
|||
orf108ng-1    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE 60

orf108-1.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT 120
|||
orf108ng-1    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT 120

orf108-1.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH 181
|||||
orf108ng-1    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH 181

```

The complete length ORF108ng nucleotide sequence (SEQ ID NO: 429) is:

```

1  ATGCTGAAAa tacctTTTGC CGTGTtgggc ggCtgccctGC TGCTTGCCGC
51  CTGCGGCAAA TCCGAAAATa cggcggaACA GCCGCAAAAT gcggCACAAA
20 101  GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ACATCGACAA TACGGCGATT
151  GCCGGTTTGG CTTTGGGACA AAGTAGCGAA GGCAAAACCA acgacgGCAA
201  AAAACAAATC AGTTATccgA TTAAAGGCTT GCCGGAACAA Aacgccgtcc
251  gGCTGACCGG AAAGCATCCC AACGACTTGG Aagccgtcgt CGGCAAATGT
301  ATGGAAACCG ACGGAAAGGA CGCGCCTTCG GGCTGGGCGG AAAACGGCGT
351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
25 401  GCAAACCTGAC TGATTACCTG ATTTTCGATT CCGCCCTGCA ACCCTATCAG
451  GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501  AATCGACAGC GagggGGCGT TTTATttccg ccgccccat tattgA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 430):

```

30 1  MLKIPFAVLG GCLLLAACGK SENTAEQPQN AAQSAPKPVF KVKYIDNTAI
51  AGLALGQSSE GKTNDGKKQI SYPIKGLPEQ NAVRLTGKHP NDLEAVVGKC
101  METDGKDAPS GWAENGVCHT LFAKLVGNI EDGGKLT DYLVSHAALQPYQ
151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 51**

The following DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 431):

```

1  ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGTTG CGATGATTGC
51  CGgATTTATC GATgcatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCC
101  CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
5  151  CTGCAAgCAG CCGCTGCTAC GTTTTTCAGCT ACGGTTTCTT TTGCACGCAA
201  AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251  TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGTTTC CAAAGATATT
301  CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
10  351  GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401  TTTTCTGTT cGGGCTGACG GTCGC .ACCG CTTTGGGTT TTTACGACGG
451  TGTGTTGCGA CCGGGTGTCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
501  TGCTCGGCTG CAAgCTGTTG AACGCGATGT CTTACACCAA ATTGGCGAAC
551  GTTGCTGCA ATCTTGTTTC GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT
15  601  TATTTTCCCG ATTGCGGCAA CGaTGGCGGT CCGTGCGTTT GTCGGtGCGA
651  ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 432; ORF109):

```

1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51  LQAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
20  101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGLRR
151  CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
201  YFPDCGNDGG RCVCRCEFRG EICRTLRFEA D*

```

Further work revealed the following DNA sequence (SEQ ID NO: 433):

```

25  1  ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGTTG CGATGATTGC
51  CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCCC
101  CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151  CTGCAAGCAG CCGCTGCTAC GTTTTTCAGCT ACGGTTTCTT TTGCACGCAA
201  AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
30  251  TAGGCGGCGT GGCCGGTGCA TTATCGGTCA GCTTGTTTC CAAAGATATT
301  CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
351  GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401  TTTTCTGTT CGGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT
451  GTGTTGCGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
35  501  GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
551  TGCCTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGAT
601  ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTG TCGGTGCGAA
651  TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
40  701  TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
751  AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 434; ORF109-1):

```

1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51  LQAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
45  101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151  VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
201  IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
251  RNPLYQMIVS MF*

```

50 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 (SEQ ID NO: 432) shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) (SEQ ID NO: 436) from strain A of *N. meningitidis*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      orf109a     MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
              10      20      30      40      50      60

10     orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      orf109a     TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
              70      80      90      100     110     120

15     orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ
      orf109a     KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
              130     140     150     160     170     180

```

20 The complete length ORF109a nucleotide sequence (SEQ ID NO: 435) is:

```

      1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGT TTGGTTG CGATGATTGC
     51  CCGATTTATC GATGCGATTG CGGGTGGGGG TGGTTTGATT ACGCTGCCTG
    101  CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
    151  CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
    201  AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCGGCA GCATCGTTTG
    251  CAGGCGGCGT GGTGCGTGCA TTATCGGTCA GCTTG GTTTC CAAAGATATT
    301  CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
    351  GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
    401  TTTTCTGTT CGGTCTGACG GTTGCAACAC TTTTGGGTTT TTACGACGGT
    30  451  GTGTTCCGAC CGGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
    501  GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
    551  TTGCCTGCAA TCTTG GTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
    601  ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTG TCGGTGCGAA
    651  TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
    35  701  TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
    751  AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 436):

```

40      1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
     51  LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
    101  LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
    151  VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGSI
    201  IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
    45  251  RNPLYQMIVS MF*

```

ORF109a (SEQ ID NO: 436) and ORF109-1 (SEQ ID NO: 434) show 99.2% identity in 262 aa overlap:

-334-

		10	20	30	40	50	60
	orf109a.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
5	orf109-1	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA					
		10	20	30	40	50	60
	orf109a.pep	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSLSLVSKDILLAVVPVLLIFVALYFVFSP					
10	orf109-1	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSLSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120
	orf109a.pep	KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK					
15	orf109-1	KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK					
		130	140	150	160	170	180
	orf109a.pep	LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI					
20	orf109-1	LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI					
		190	200	210	220	230	240
	orf109a.pep	SMAVKLLIDERNPLYQMIVSMFX					
25	orf109-1	SMAVKLLIDERNPLYQMIVSMFX					
		250	260				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 (SEQ ID NO: 432) shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) (SEQ ID NO: 438) from *N. gonorrhoeae*:

30	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
	orf109ng	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
	orf109.pep	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSLSLVSKDILLAVVPVLLIFVALYFVFSP	120
	orf109ng	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSLSLVSKDILLAVVPVLLIFVALYFVFSP	120
35	orf109.pep	KLDGSKEGKARMSFFLFGLTVXTAFGLRRCVRTGCRLVFSDCLYCFARLQAUERDVLHQ	180
	orf109ng	KLDGSKEGKARMSFFLFGLTVATAFGFLRRCVRTGCRLVFSDCLYCFARLQAUERDVLHQ	180
	orf109.pep	IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRTLRFED	231
40	orf109ng	IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFED	231

An ORF109ng nucleotide sequence (SEQ ID NO: 437) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 438):

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK  
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI  
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR  
 151 CVRTGCRLVF SDCLYCFARL QAUERDVLHQ IGERCLQSWF AIGIPAARFD  
 201 YFPDCGNDGG RCVCRCEFR C EICRPLRFEA D\*

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 439):

1 ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATCGC  
 51 CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG  
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG  
 151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA  
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG  
 251 CAGGCGGCGT GGTCCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT  
 301 TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT  
 15 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT  
 401 TTTTCTATT CGGGCTGACG GTTGCACCGC TTTTGGGTTT TTACGACGGT  
 451 GTGTTCCGAC CGGGTGTCCG CTCGTTTTTT CTGATTGCCT TTATTGTTTT  
 501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG  
 551 TTGCTTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT  
 20 601 ATTTTCCCGA TTGTGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA  
 651 TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC  
 701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG  
 751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTAA

25 This corresponds to the amino acid sequence (SEQ ID NO: 440; ORF109ng-1):

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK  
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI  
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG  
 151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I  
 201 IFPIVATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE  
 251 RNPLYQMIVS MF\*

ORF109ng-1 (SEQ ID NO: 440) and ORF109-1 (SEQ ID NO: 434) show 98.9% identity in 262 aa overlap:

35 orf109ng-1.pep 10 20 30 40 50 60  
 MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  
 orf109-1 MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  
 10 20 30 40 50 60  
 40 orf109ng-1.pep 70 80 90 100 110 120  
 TVSFARKGLIDWKKGLPIAAASFAGGVVVALSVSLVSKDILLAVVPVLLIFVALYFVFSP  
 orf109-1 TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  
 70 80 90 100 110 120  
 45 orf109ng-1.pep 130 140 150 160 170 180  
 KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK  
 orf109-1 KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK  
 130 140 150 160 170 180  
 50 orf109ng-1.pep 190 200 210 220 230 240

-336-

```

orf109ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf109-1        LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
                190      200      210      220      230      240

5              250      260
orf109ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
                ||||||||||||||||||||||||||||||||||
orf109-1        SMAVKLLIDERNPLYQMIVSMFX
                250      260

```

In addition, ORF109ng-1 (SEQ ID NO: 440) shows homology to a hypothetical *Pseudomonas* protein (SEQ ID NO: 1140):

```

sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3' REGION (ORF9)
)gi|94984|pir|I38164 hypothetical protein 9 - Pseudomonas sp )gi|551929 (M62866)
ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXXXXXKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
          PP+  + TNKLQ                      R+G ++ K+ LP+                      D+
Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPLKDGSKGKARMSFFLFGTLVAPLLGFYDGVFGPGVGSFF 160
          L A++P LLI +ALYF  P + G +  +R++ F+F LT+ PL+GFYDGVFGPG GSFF
Sbjct: 103 LKAILPFLIIAIALYFGLKPNM-GDVDQHSRVTPFVFTLTLVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMSYSYTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
          ++ F+ L G +L A ++TK N  N+G+ VFL G++++ +  M +G F+GA +G+
Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLIVISISMAVKLLIDERNPL 254
          R+A+  G+K+IKPLL+++SI++A++LL D  +PL
Sbjct: 222 RYAMAKGAKIIKPLLIVISIALAIRLLADPTHPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 52

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 441):

```

1  ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
51  CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTTGGG CGAG.ATTTT
101 TGGTTTTCTG GGA CTGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
151 TCATGATGTT TTTGGTGGTT TCTACCAGTT TGTGCCTGAT TCGCAATGTG
201 CCGCCGTTCT GCGCGGAAAT GAAGTCTTTT CGGGAAAAGG TTAAAGAAAA
251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGGATGTA AAAATTGCGC

```

301 CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTC A GGGGAAAACC  
 351 ATTAACCGTG AAGACGGGTC GGTTCGTGATT GCCGCCAAAA AAGGCACAAT  
 401 GAACAAATGG GGCTATATCT TTGCCCATGT TGCTTTGATT GTCATTTGCC  
 451 TGGGCGGGTT GATAGACAGT AACCTGCTGT TGAAACTGGG TATGCTGACC  
 501 GGTTCGGATTG TTCCGACAA TCAGGCGGTT TATGCCAAGG ATTTC.AAGC  
 551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA  
 601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence (SEQ ID NO: 442; ORF110):

1 .LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI  
 51 MMFLVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP  
 101 EVAKRYLEVQ GFQGTINRE DGSVLIAAKK GTMNKWGYIF AHVALIVICL  
 151 GGLIDSNLLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI\*GQRQY  
 201 FXRGRVRMWF S\*

Computer analysis of this amino acid sequence gave the following results:

Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 (SEQ ID NO: 442) shows 91.5% identity over a 188aa overlap with ORF88a (SEQ ID NO: 332) from strain A of *N. meningitidis*:

20		10	20	30	40	50	60
	orf88a.pep	MSKSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA					
	orf110						
				10	20	30	
25		70	80	90	100	110	120
	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
	orf110	XIFGFLGLYDVYASAWFVIMMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
		40	50	60	70	80	90
30		130	140	150	160	170	180
	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL					
	orf110	SSLLDVKIAPEVAKRYLEVQGFQGTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL					
		100	110	120	130	140	150
35		190	200	210	220	230	240
	orf88a.pep	GGLIDSNLLLKLGMALTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF					
	orf110	GGLIDSNLLLKLGMALTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVRMWF					
		160	170	180	190	200	210
40		250	260	270	280	290	300
	orf88a.pep	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT					
	orf110	SX					

However, ORF88 (SEQ ID NO: 328) and ORF110 (SEQ ID NO: 442) do not align, because they represent two different fragments of the same protein.

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 (SEQ ID NO: 442) shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) (SEQ ID NO: 444) from *N. gonorrhoeae*:

	orf110.pep	LLGIASVIGTLLQQNQPTDYLVKFGSFWA	30
		:	
	orf110ng	MSKSRIPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLVKFGPFWT	60
10	orf110.pep	XIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGTINREDGSLIAAKKGTMNKWDYIFAHVALIVICL	150
		:       :	
	orf110ng	SSLLDVKIAPEVAKRYLEVRGFQGTIVSREDGSLIAAKKGTMNKWDYIXAHVALIVICL	180
15	orf110.pep	GGLIDSNLLKLGLMTGRIFRTIRRFMPRIKXKPEXFGCVQSLIXGQRQYFXRGRVWMWF	210
		:      :     :             :	
	orf110ng	GRLINXNLLKLGLMAGSIFRNNRRVMPRIKXKPEXFGCVQSLIXGQRQYFQGRVWMWF	240
	orf110.pep	S 211	
20			
	orf110ng	S 241	

The complete length ORF110ng nucleotide sequence (SEQ ID NO: 443) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 444):

25	1	MSKSRIPTL	LSRPWFAFFS	SMRFAVALLS	LLGIASVIGT	VLQQNQPTD
	51	YLVKFGPFWT	RIFDFLGLYD	VYASAWFVVI	MMFLVVSTSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SSLLDVKIAP	EVAKRYLEVR	GFQGTVSRE
	151	DGSLVIAAKK	GTMNKWDYIX	AHVALIVICL	GRLINXNLLL	KLGLMAGSIF
	201	RNNRRVMPRI	SKPESIWGGV	QSLIKGQRQY	FQGRVWMWF	S*

Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### **Example 53**

The following DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 445):

35	1	ATGCCGTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGTCT	TGATATTTGC
	51	CCTGGGTTTC	ATCTTCCTGA	ACGCCTGTTC	GGAACAAACC	GCGCAAACCG



5  
10  
15  
20

```

101 TTACCCTGCA AGGCGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT
151 TCAAATAATC GGGACAAACT CCCCTCACCT GCCGAAATAC AAAAACGCAT
201 CGATGACGCG CTTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
251 ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCCTCCGC
301 ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC GCCTGAACCG
351 CCTGACACAC GGC GCGCTGG ACGTAACCGT CGGCCCTTG GTCAACCTTT
401 GGGGATTTCG CCCCACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA
451 ATCAAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTTGAAACA
501 AGGCAAAGAT TACGCTTCCT TGAGCAAAAC CCACCCCAAG GCCTATTTCG
551 ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAAGT TGCGGGCGAA
601 CTGGAAAAAT ACGGCATTCA AAATTATCTG GTCGAAATCG GCGGCGAGTT
651 GCACGGCAAA GGCAAAAACG CGCGCGGCGA ACCGTGGCGC ATCGGTATCG
701 AGCAGCCCAA TATCGTCCAA GCGGCAATA CGCAGATTAT CGTCCCCTG
751 AACCAACGTT CGCTTGCCAC TTCCGGCGAT TACCGTATTT TCCACGTCGA
801 TAAAAACGGC AAACGCCTCT CCCATATCAT CAACCCGAAC AACAAACGAC
851 CCATCAGCCA CAACCTCGCC TCCATCAGCG TGGTCGAGA CAGTGCGATG
901 ACGGCGGACG GCTTGTCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
951 CTTAAAGCTG GCAGAGCGCG AAAAATCGC TGTTTTCTG ATTGTCAGGG
1001 ATAAAGGCG CTACCGCACC GCCATGTCTT CCGAATTGA AAAACTGCTC
1051 CGCTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 446; ORF111):

25  
30

```

1 MPSETRLPNF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTTYTVKYL
51 SNNRDKLPSP AEIQKRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR
101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFDPK SVTREPSPEQ
151 IKQAASYTGI DKIIILKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
201 LEKYGIQNYL VEIGGELHGK GKNARGEPR IGIEQPNIVQ GGNTQIIIVPL
251 NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVDASAM
301 TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
351 R*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 (SEQ ID NO: 446) shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) (SEQ ID NO: 448) from strain A of *N. meningitidis*:

35  
40  
45

```

          10      20      30      40      50      60
orf111a.pep MPSETRLPNFIRTLIFALSFI FLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDXLPSP
          |||||:||||:|||||
orf111      MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSP
          10      20      30      40      50      60

          70      80      90      100     110     120
orf111a.pep AEIQXRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTA EAVHLNRLTH
          ||| |||||:|||||
orf111      AEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH
          70      80      90      100     110     120

          130     140     150     160     170     180
orf111a.pep GALDVTVGPLVNLWGFDPKSVTREPSPEQIKQAASYTGIDKIILKQKGDYASLSKTHPK
          |||||:|||||
orf111      GALDVTVGPLVNLWGFDPKSVTREPSPEQIKQAASYTGIDKIILKQKGDYASLSKTHPK
          130     140     150     160     170     180

```

-340-

		190	200	210	220	230	240
	orf111a.pep	AYLDLSSIAGFGVDXVAGELEKYGIQNYLVEIGGELHGKXKNARGEPWRIGIEQPNIVQ					
	orf111	AYLDLSSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGKNARGEPWRIGIEQPNIVQ					
5		190	200	210	220	230	240
	orf111a.pep	250	260	270	280	290	300
	orf111	GGNTQIIIVPLNNRSXATSGDYRIFHVDKSGKRLSHIINPNNKRPISHNLAISISVXADSAM					
10		250	260	270	280	290	300
	orf111a.pep	310	320	330	340	350	
	orf111	TADGXSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
15		310	320	330	340	350	

The complete length ORF111a nucleotide sequence (SEQ ID NO: 447) is:

	1	ATGCCGTC	TAAGACGC	CTGCCGAAC	TTTATCCGC	ACCTTGAT	ATTTTGC
	51	CCTGAGTT	TTATCTTC	TGACGCT	GTACGGA	AAACCGC	CGGCAACC
20	101	TTACCCTG	CAAGGTGA	AACATGG	GCACCTA	TCCTACCT	CAAAACCT
	151	TCAAATAA	TCGGGAC	NAACCCC	NTCACCT	GCCGAAAT	ACAAAACG
	201	CGATGACG	CGCTTAA	AAGTCA	ACCGGCA	GATGTCC	ACCATCAG
	251	ACTCCGAA	ATCAGCC	GTTC	AACCAAC	ACAGCCG	CAAGCCCT
	301	ATTTCAAG	CGACTTC	GCACA	CGTTACT	GCC	GAAGCCGT
25	351	CCTGACAC	ACGGCGC	CTGGAC	GTAACCG	TGCGCCCT	GTCAACCT
	401	GGGGATT	TCGGCCC	GACAAA	ATCCGTT	ACCGT	CGCCGAAC
	451	ATCAAACA	AGCAGCA	TCTTA	TACGGGC	ATACAAA	ATCAATTT
	501	AGGCAAG	ATACGCT	TCCTTG	AGCAAA	ACCCCA	AGGCCTAT
	551	ATTTATCT	TTCGAT	TCCAAA	GGCTTC	CGGCG	TTGATNAN
30	601	CTGGAATA	ATACGGC	ATTCA	TAATAT	CTGT	CGGAAAT
	651	GCACGGCA	AAAGNCA	AAACG	CGCGCG	CGACCT	TGGCGC
	701	AACAGCC	CAATCG	TCCAA	GGCGCA	ATACG	AGATTAT
	751	AACAACCG	TTCGNT	TGCCAC	TTCCGG	CGAT	TACCGTAT
	801	TAAAAGCG	GCAACG	CTCT	CCCATAT	CAAT	CCGAAC
35	851	CCATCAG	CAACCT	TCGCC	TCCATC	AGCG	TGNTCGCA
	901	ACGGCGG	ACGCTT	NTCCAC	AGGATT	ATTC	GTATTGGG
	951	CTTAAAG	CTGAG	CGCG	AAAACT	TCGC	TGTTTTC
	1001	ATAAAGC	GGCTAC	CGCACC	GCCATG	TCTT	CCGAATTG
40	1051	CGCTAA					

This encodes a protein having amino acid sequence (SEQ ID NO: 448):

	1	MPSETRLPNF	IRTLIFALSF	IFLNACSEQT	AQTVTLQGET	MGTTYTVKYL
	51	SNNRDXLPSP	AEIQXRIDDA	LKEVNRQMST	YQPDSEISRF	NQHTAGKPLR
45	101	ISSDFAHVTA	EAVHLNRLTH	GALDVTVGPL	VNLWGFGPDK	SVTREPSPEQ
	151	IKQAASYTGI	DKIILKQKQD	YASLSKTHPK	AYLDLSSIAK	GFGVDXVAGE
	201	LEKYGIQNYL	VEIGGELHGK	XKNARGEPWR	IGIEQPNIVQ	GGNTQIIIVPL
	251	NNRSXATSGD	YRIFHVDKSG	KRLSHIINPN	NKRPISHNLA	SISVXADSAM
	301	TADGXSTGLF	VLGETEALKL	AEREKLAVFL	IVRDKGGYRT	AMSSEFEKLL
	351	R*				

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 (SEQ ID NO: 446) shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) (SEQ ID NO: 450) from *N. gonorrhoeae*:

5	orf111ng	10 20 30 40 50 60	MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPS
	orf111	10 20 30 40 50 60	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPS
10	orf111	70 80 90 100 110 120	AKIQKRIDDALKEVNRQMSTYQTDSEISRFNQHTAGKPLRISSDFAHVTAEA VRLNRLTH
	orf111	70 80 90 100 110 120	AEIQKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTAEA VRLNRLTH
15	orf111ng	130 140 150 160 170 180	GALDVTVGPLVNLWGFGDPKSVTREPSPEQIKQAASYTGIDKII LQQGKDYASLSKTHPK
	orf111	130 140 150 160 170 180	GALDVTVGPLVNLWGFGDPKSVTREPSPEQIKQAASYTGIDKII LKQGKDYASLSKTHPK
20	orf111ng	190 200 210 220 230 240	AYLDLSSIAGFGVDK VAGELEKYGIQNYLVEIGGELHGKGKNAHGEPWRIGIEQPNIIQ
	orf111	190 200 210 220 230 240	AYLDLSSIAGFGVDK VAGELEKYGIQNYLVEIGGELHGKGKNAHGEPWRIGIEQPNIVQ
25	orf111ng	250 260 270 280 290 300	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLA SISVVS DSAM
	orf111	250 260 270 280 290 300	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLA SISVVA DSAM
30	orf111ng	310 320 330 340 350	TADGLSTGLFVLGETEALRLAEQEKLA VFLIVRDKDGYRTAMSSEFAKLLRX
	orf111	310 320 330 340 350	TADGLSTGLFVLGETEAL KLAEREKLA VFLIVRDKGGYRTAMSSEFEKLLRX

The complete length ORF111ng nucleotide sequence (SEQ ID NO: 449) is:

35	1	ATGCCGTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTTC	ATCTTCCTGA	ACGCCTGTTC	GGaacaacC	GCGCAaaccg
40	101	TTACCCTGCA	AGGCGAAAcg	aTGGGTACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACAAACT	CCCCTCCCT	GCCAAAATAC	AAAAGCGCAT
45	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TACCAGACCG
	251	ATTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCTCCGC
50	301	ATTTCAAGCG	ATTTCCGCACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
	351	CCTGACTCAC	GGCGCACTGG	ACGTAACCGT	CGGCCCTTTG	GTCAACCTTT
55	401	GGGGGTTTCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTGTCAACA
60	501	AGGCAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
65	601	CTGGA AAAAT	ACGGCATTCA	AAATTATCTG	GTCGAAAcg	gcggcGAGTT
	651	GCACGGCAAA	GGCAAAAATG	CGCACGGCGA	ACCGTGCGCG	ATCGGTATAG
70	701	AGCAACCCAA	TATCATCCAA	GgcgGCAata	CGCAGATTAt	cgtcccgtcg
	751	aaCaaccgtt	cgctTGCCAC	TTCCGGCGAT	TAcgtaTTT	tccacgtcgA

-342-

5  
 801 TAAAAAcggc aaacgccttt cccacaTCAT CAATCCCaAC aacAAACgac  
 851 ccATCAGcca caacctcgcc tccatcagcg tggctctcAGA CAGTGCAATG  
 901 ACGGCGGACG GTTtatCCAC AGGATTATTT GTTTTAGGCG AAACCGAAGC  
 951 CTTAAGGCTG GCAGAACAAAG AAAAActCGC TGTTTTCCTA ATTGTCCGGG  
 1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTTCG CAAGCTGCTC  
 1051 CGCTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 450):

10  
 1 MPSETRLPNL IRALIFALGF IFLNACSEQT AQTVTLQGET MGTTYTVKYL  
 51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMS YQTDSEISRF NQHTAGKPLR  
 101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFGPDK SVTREPSPEQ  
 151 IKQAASYTGI DKIIQQGKD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE  
 201 LEKYGIQNYL VEIGGELHGK GKNAHGEPWR IGIEQPNIIQ GGNTQIIIVPL  
 251 NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVSDSAM  
 15 301 TADGLSTGLF VLGETEALRL AEQEKLAVFL IVRDKDGYRT AMSSEFAKLL  
 351 R\*

This protein shows homology with a hypothetical lipoprotein precursor (SEQ ID NO: 1141) from *H. influenzae*:

20 sp|P44550|YOJL\_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR )gi|1074292|pir|4  
 hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20) )gi|1573128  
 (U32702) hypothetical [Haemophilus influenzae] Length = 346  
 Score = 353 bits (896), Expect = 9e-97  
 Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)

25 Query: 7 LPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGTTYTVKYLSNNRDKLPSPAKIQKR 66  
 + LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +  
 Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTTYHVKYLDGGSITATSE-KTHEE 58

30 Query: 67 IDDALKEVNRQMSYQTDSEISRFNQHT-AGKPLRISSDFAHVTA EAVRLNRLTHGALDV 125  
 I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV  
 Sbjct: 59 IEAILKDVNAKMSTYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTGALDV 118

Query: 126 TVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIIQQGKDYASLSKTHPKAYLDL 185  
 TVGP+VNLWGFGP+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL  
 Sbjct: 119 TVGPVNLWGFGPEKRPEKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178

35 Query: 186 SSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGKNAHGEPWRIGIEQPNIIQGGNTQ 245  
 SSIAGFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +

Sbjct: 179 SSIAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKPWQIAIEKPTTTGERAVE 238  
 Query: 246 IIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLASISVVSDSAMTADGL 305  
 ++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL  
 Sbjct: 239 AVIGLNNMGMASGDYRIY-FEENGKRAFEIDPKTGYPHQHHLASITVLAPTSMTADGL 297

40 Query: 306 STGLFVLGETEALRLAEQEKLAVFLIVRDKDGYRTAMSSEFAKL 349  
 STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL  
 Sbjct: 298 STGLFVLGEDKALEVAEKNLAVYLIIRTDNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 45 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 54**

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 451):

```

1  ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCGTCA
5  51  AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG
101 GCGGCGCGGC TGGCGACGGG TGGCGCAAAG GCGTGCAAAT CCGGCGGCGAG
151 GTGTTTGTAC GGCAAAATGA AGGCAGCCkA yTGGCAATCG GCGTGATGGG
201 CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
251 gCAGTGATTT GTATGGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
301 TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
10 351 ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG
451 GAAGGCATTG TCGGAAAAGG CAATAATGTG CCGTTTTACC TACAACCGCA
501 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTTACC GACAGCGAGG
551 GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC
15 601 AtTCGGGCAA AAACCCGTTT TGCTTTGCGT AACGGTGTCA ATCTTCAGCC
651 TTTTGCCGCT TTTAATGTtT TGCACAGGTC AAAATCTTTC GGCGTGGA
701 TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG
751 TTCGGTATTG AAGCCGTTG GAAAGGCCAT ATGTCCGCA..

```

20 This corresponds to the amino acid sequence (SEQ ID NO: 452; ORF35):

```

1  ..PCRRQDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGVQIGGE
51 VFVRQNEGSX LAIGVMGGRA GQHASVNGKG GAAGSDLYGY GGGVYAAWHQ
101 LRDQKTGAYL DGWLQYQRFK HRINDENRAE RYKTKGTAS VEGGYNALVA
151 EGIVGKGNV RFYLQPPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
201 IRAKTRFALR NGVNLQPFPA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
251 FGIEAGWKGH MSA..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

30 ORF (SEQ ID NO: 452) and virg-h protein (SEQ ID NO: 1146) show 51% aa identity in 261aa overlap:

```

Orf35 5  QGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKGVQIGGEVFVRQNEGSXLAI 63
      + D++      R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I
virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQWVQGKTAPVEGYRKGVLGGEVFTWQNESNQLSI 455

35  Orf35 64  GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDQKTGAYLDGWLQYQRFKH 121
      G+MGG+A Q ++ +      ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H
virg-h 456 GLMGGQAEQRSTFHNPDTDNLTGTVNGKFGAGVYATWHQLQDKQTGAYADSWMQYQRFH 515

Orf35 122 RINDENRAERYKTKGTASVEGGYNALVAEGIVGKGNVRFYLQPPAQFTYLGVNGGFTD 181
      RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLQPPAQ TYLGVNG F+D
40  virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTTKGNSLRVYLQPPAQLTYLGVNGKFS 575

Orf35 182 SEGTAVGLLGSQWQSRAGIRAKTRFALRNGVNLQPFPAFNVLHRSKSFVEMDGEKQTL 241
      SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVMEDGE++ +
virg-h 576 SENAHVNLLGSRQLQTRVGVOAKAQFSLYKNIAIEPFAAVNALYHNKPFVEMDGERVI 635

```

Orf35 242 AGRTALEGRFGIEAGWKGHMS 262  
+TA+E + G+ K H++  
virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

### Homology with a predicted ORF from *N.meningitidis* (strain A)

- 5 ORF35 (SEQ ID NO: 452) shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) (SEQ ID NO: 454) from strain A of *N. meningitidis*:

	orf35.pep		10	20	30
			PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIRG		
10	orf35a	QRLAIPAEAEVLYAQQAYAANTLFLGLRAADRGDDVYAADPSRQKLWLRFIGGRSHQNIRG			
		310 320 330 340 350 360			
	orf35.pep		40	50	60
			GAAADGWRKGVQIGGEVFRQNEGSXLAIGVMGGRAGQHASVNGKGAAGSDLYGYGGGV		
15	orf35a	GAAADGRRKGVQIGGEVFRQNEGSRLAIGVMGGRAGQHASVNGKGAAGSYLHGYGGGV			
		370 380 390 400 410 420			
	orf35.pep		100	110	120
			YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV		
20	orf35a	YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV			
		430 440 450 460 470 480			
	orf35.pep		160	170	180
			GKGNNVRFYLPQAQFTYLGVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVN		
25	orf35a	GKGNNVRFYLPQAQFTYLGVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRNGVN			
		490 500 510 520 530 540			
	orf35.pep		220	230	240
			LQPF AAFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA		
30	orf35a	LQPF AAFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSARIGYGKRTDGD			
		550 560 570 580 590 600			
	orf35a	KEAALSLKWLFX			
		610 620			

- 35 The complete length ORF35a nucleotide sequence (SEQ ID NO: 453) is:

	1	ATGTT	CAGAG	CTCAG	CTTGG	TTCAA	ATACT	CGTT	CTACCA	AAAT	CGGCGA
	51	CGAT	GCCGAT	TTTTC	ATTTT	CAGAC	AAGCC	GAAAC	CCGGC	ACTT	CCCATT
	101	ATTTT	TCCAG	CGGT	AAAACC	GATCA	AAATT	CATCC	GAATA	TGGG	TATGAC
40	151	GAAAT	CAATA	TCCA	AGGTAA	AAACT	ACAAT	AGCGG	CATAC	TCGCC	GTCGA
	201	TAAT	TGCCCC	GTTGT	TAAAGA	AATAT	ATTAC	AGATA	CTTAC	GGGG	GATAATT
	251	TAAAG	GATGC	GGTTA	AAGAAG	CAATT	ACAGG	ATTTA	TACAA	AACA	AGACCC
	301	GAAG	CTTGGG	AAGAA	AATAA	AAAAC	GGAAT	GAGGA	GGCGT	ATATA	GAAACA
	351	GCTT	GGACCA	AAATT	TAGTA	TACTC	AAACA	GAAAA	AACCC	GATTT	AATTA
	401	ATAAA	TGGT	AGAAG	ATTCC	GTA	CTCACTC	CTCAT	AGTAA	TACAT	CACAG
45	451	ACTAG	TCTCA	ACAAC	ATCTT	CAATA	AAAAAA	TTACA	CGTCA	AAAT	CGAAAA
	501	CAAAT	CCCCAC	GTCGC	CGGAC	AGGT	GTTGGA	ACTGA	CCAAG	ATGAC	GCTGA

5 551 AAGATTCCTT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA  
 601 ACTTCCGATA ATGCCCGCAT CCGCCTGAAC ACGAAAGATG AAAAAGTAC  
 651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCCTGTTC GGCTACGACG  
 701 TGCGGGAGTC GGACAAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA  
 751 CAATCCGGCG TGGTTTTTGA ACGCCGCGCG GAAAACTGA AAACGCTCGA  
 801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTTGCCT  
 851 TTAAACAAAA TTACCGGCAG GGAAGTACG AATTATTGCT CAAGCAATGC  
 901 GAAGGCGGAT TTTGCTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA  
 951 AGCGGTTTTA TATGCCAAC AGGCTTATGC GGCAATACT TTGTTCGGG  
 10 1001 TGCGTGCCGC CGACAGGGG GACGACGTGT ATGCCCGCGA TCCGTCCCGT  
 1051 CAAAAATTGT GGCTGCGCTT CATCGGCGGC CGGTGCGATC AAAATATACG  
 1101 GGGCGGCGCG GCTGCGGACG GCGGCGCAA AGGCGTGCAA ATCGGCGGCG  
 1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGCAAT CCGCGTGATG  
 1201 GCGGCGAGG CTGGCCAGCA CGCATCAGTC AACGGCAAAG GCGGTGCGGC  
 15 1251 AGGCAGTTAT TTGCATGGT ATGGCGGGG TGTATTATGT CCGTGGCATC  
 1301 AGTTGCGCGA TAAACAAAC GGTGCGTATT TGGACGGCTG GTTGCAATAC  
 1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA  
 1401 AACCAAAGGT TGGACGGCTT CTGTGCAAGG CCGCTACAAC GCGCTGTGG  
 1451 CGGAAGGCGT TGTCGGAATA GGCAATAATG TGCGGTTTTA CCTGCAACCG  
 20 1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GCGGCTTTA CCGACAGCGA  
 1551 GGGGACGGCG GTCGGACTGC TCGGCAGCGG TCAGTGGCAA AGCCGCGCGG  
 1601 GCATTCGGGC AAAAACCCTT TTTGCTTTC GTAACGGTGT CAATCTTCAG  
 1651 CCTTTTGCCG CTTTTAATGT TTTGCACAGG TCAAAATCTT TCGGCGTGA  
 25 1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC  
 1751 GGTTCGGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACGCATCGGA  
 1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGCATTGT CGCTCAAATG  
 1851 GCTGTTTTGA

This encodes a protein having amino acid sequence (SEQ ID NO: 454):

30 1 MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT DQNSSEYGYD  
 51 EINIQGNYN SGILAVDNMP VVKYITDLY GDNLDKAVKK QLQDLYKTRP  
 101 EAWEEKKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ  
 151 TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE  
 201 TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESKDP ALTFEEKVSG  
 35 251 QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLKQC  
 301 EGGFCLGVQR LAIPEAEAVL YAQQAYAAANT LFGLRAADRQ DDVYAADPSR  
 351 QKLWLRFIGG RSHQNIIRGGA AADGRRKGVQ IGGEVFVRQN EGSRLAIGVM  
 401 GGRAGQHASV NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLQY  
 451 QRFKHRINDE NRAERYKTG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ  
 40 501 QAQFTYLVGN GGFTDSEGTA VGLLGSQWQ SRAGIRAKTR FALRNGVNLQ  
 551 PFAAFNVLHR SKSFGVEMDG EKQTLAGRTA LEGRFGIEAG WKGHMSARIG  
 601 YGKRTDGDKE AALSLKWL\*  
 601

### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF35 (SEQ ID NO: 452) shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) (SEQ ID NO: 456) from *N. gonorrhoeae*:

orf35.pep PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIIRG 34  
 ::||:: ||||| ||:|::| ::|  
 orf35ngh FTKVQERDDIAIYAQQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370  
 50 orf35.pep GAA-ADGWRKGVQIGGEVFVRQNEGSXLAIGVMGGRAGQHASVNGKG--GAAGSDLYGYG 91  
 :| ::||:||||:||||: ||::| :||:||||: |::: : : : : :|  
 orf35ngh KTAPEGYRKGVLGGEVFTWQNESNQSLIGLMGGQAEQRSTFRNPDTDNLTGTVKGF 430

5	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTGWASVEGGYNALVAE	151
	orf35ngh	AGVYATWHQLQDKQTGAYVDSWMQYQRFHRINTEYATERFTSKGITASIEAGYNALLAE	490
	orf35.pep	GIVGKGNNVRFYLPQAQFTYLVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRN	211
	orf35ngh	HFTKKGNSLRVYLQPAQLTYLVNGKFSSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
	orf35.pep	GVNLQPPFAAFNVLHRSKSFGEVMDGEKQTLAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	GVTFPFPFVAVNSIYQQKPFGEIDGDRRVINNKTVIETQLGVAAKIKSHLTLQASFNRQT	610

A partial ORF35ngh nucleotide sequence (SEQ ID NO: 455) is predicted to encode a protein having partial amino acid sequence (SEQ ID NO: 456):

15	1	..KKLRDRNSEY WKEETYHIKS NGRTPNIPA LFPKHPFDPF ENINNSKKIS
	51	FYDKEYTEDY LVGFARGFGV EKRNGEEKP LRQYFKDCVN TENSNNNDNCK
	101	ISSFGNYGPI LIKSDIFALA SQIKNSHINS EILSVGNIE WLRPTLNKLT
	151	GWQEHLIYAGL DPFHYIEVTD NSHVIGQTID LGALELTNSL WKPRWNSNID
	201	YLITKNAEIR FNTKNESLLV KEDYAGGARF RFAYDLKDKV PEIPVLTFEK
	251	NITGTSDIIF EGKALDNLKH LDGHQIVKVN DTADKDAFRL SSKYRKGITYT
20	301	LSLQQRPEGF FTKVQERDDI AIYAQQAQAA NTLFALRLND KNSDIFDRTL
	351	PRKGLWLRVI DGHSNQWVQG KTAPVEGYRK GVQLGGEVFT WQNESNQLSI
	401	GLMGQAQEQR STFRNPDTDN LTGNVKGFG AGVYATWHQL QDKQTGAYVD
	451	SWMQYQFRFH RINTEYATER FTSKGITASI EAGYNALLAE HFTKKGNSLR
	501	VYLQPAQLT YLVNGKFS SENAQVNLLG SRQLQSRGV QAKAQFAFTN
	551	GVTFPFPFVAV NSIYQQKPFGEIDGDRRVI NNKTVIETQL GVAKIKSHL
25	601	TLQASFNRQT SKHHHAKQGA LNLQWTF*

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 55

30 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 457):

35	1	..GCGGAATATG TTCAGTTCTC TATAGATTTG TTCAGTGTGG GTAAATCGGG
	51	GGGCGGTATA CCTAAGGCTA AGCCTGTGTT TGATGCGAAA CCGAGATGGG
	101	AGGTTGATAG GAAGCTTAAT AAATTGACAA CTCGTGAGCA GGTGGAGAAA
	151	AATGTTCAGG AAACGAGAAG AAGGAGTCAG AGTAGTCAGT TTAAAGCCCA
	201	TGCGCAACGA GAATGGGAAA ATAAAACAGG GTTAGATTTT AATCATTTTA
	251	TAGGTGGTGA TATCAATAAA AAAGGCACAG TAACAGGAGG GCATAGTCTA
40	301	ACCCGTGGTG ATGTACGGGT GATACAACAA ACCTCGGCAC CTGATAAACA
	351	TGGGT.TTA TCAAGCGACA GTGGAATTN A

40 This corresponds to the amino acid sequence (SEQ ID NO: 458; ORF46):

45	1	..AEYVQFSIDL FSVGKSGGGI PKAKPVFDAK PRWEVDRKLN KLTTRQVEK
	51	NVQETRRRSQ SSQFKAHAQR EWENKTGLDF NHFIGGDINK KGTVTGGHSL
	101	TRGDVRVIQQ TSAPDKHGXL SSDSGNX

Further work revealed further partial nucleotide sequence (SEQ ID NO: 459):



```

      1  ..GCAGTGTGCC TnCCGATGCA TGCACACGCC TCAnATTGG CAAACGATTC
    51  TTTTATCCGG CAGGTTCCTCG ACCGTCAGCA TTTCGAACCC GACGGGAAAT
  101  ACCACCTATT CGGCAGCAGG GGGGAACCTG CCGAGCGCCA GTCTCATATC
  151  GGATTGGGAA AAATACAAAG CCATCAGTTG GGCAACCTGA TGATTCAACA
    5  201  GGC GGCCATT AAAGGAAATA TCGGCTACAT TGTCCGCTTT TCCGATCACG
      251  GGCACGAAGT CCATTCCCCs TTCGACAACC ATGCCTCACA TTCCGATTCT
      301  GATGAAGCCG GTAGTCCCGT TGACGGATTT AGCCTTTACC GCATCCATTG
      351  GGACGGATAC GAACACCATC CCGCCGACGG CTATGACGGG CCACAGGGCG
      401  GCGGCTATCC CGTCCCAAA GCGCGAGGG ATATATACAG TTACGACATA
    10  451  AAAGCGTTG CCCAAAATAT CCGCTCAAC CTGACCGACA ACCGAGCAC
      501  CGGACAACGG CTTGCCGACC GTTTCACAA TGCCGGTAGT ATGCTGACGC
      551  AAGGAGTAGG CGACGATTTC AAACGCGCCA CCCGATACAG CCCCAGAGCTG
      601  GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT
    15  651  TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

```

This corresponds to the amino acid sequence (SEQ ID NO: 460; ORF46-1):

```

      1  ..AVCLPMHAHA SXLANDSFIR QVLD RQH FEP DGKYHLFGSR GELAERQSHI
    51  GLGKIQSHQL GNLMIQAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS
  101  DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PQGGGYPA PK GARDIYSYDI
  151  KGV AQNIRLN LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSP EL
    20  201  DRSGNAAEAF NGTADIVKNI IGAAGEI

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF46 (SEQ ID NO: 458) shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) (SEQ ID NO: 462) from *N. gonorrhoeae*:

```

  orf46.pep                      AEYVQFSIDLFSVGKSGGGIPKAKPVFDAKPRWEVD RKL NKL TTR      45
                                     |||
  orf46ng          PKTGV PFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVD RKL NKL TTR      217
                                     |||
  30  orf46.pep          EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHF IGGDINKKGT V TGGHSLTRGDV      105
                                     |||
  orf46ng          EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHF IGGDINKKGA V TGGHSLTRGDV      277
                                     |||

  orf46.pep          RVIQQTSA PDKHGXLSSDSGN      126
                                     |||
  35  orf46ng          RVIQQTSA PDKHGVLSSDSGN      298

```

A partial ORF46ng nucleotide sequence (SEQ ID NO: 461) is predicted to encode a protein having partial amino acid sequence (SEQ ID NO: 462):

```

    40  1  ..RRLKHCCHAR LGS AFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPQC
      51  RTRHRSRQQY LYGSHPHQRD WSCPGKIQLG RHHGTSCRAV ADXRDRICER
    101  EIRRQRQXCR CRLGKIPSL S IPKYPLKLEQ RYGKENITSS TVPPSNGK NV
    151  KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD
    201  AKPRWEVDRK LNKLT TREQV EKNVQETRRR SQSSQFKAHA QREWENKTGL
    251  DFNHFIGGDI NKKGA V TGGH SLTRGDV RVI QQTSA PDKHG VLSSDSGN*
    45

```

Further work revealed the complete gonococcal DNA sequence (SEQ ID NO: 463):

```

1  TTGGGCATTT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
51  CCTGCCGATG CATGCACACG CCTCAGATTT GGcaAACGAT CCCTTTATCC
101 GgCaggttct CGaccGTCAG CATTTcGaac ccgacggGAa ATACCaCCTA
5  151  TtcggCaGCA GGGGGGAGCT TgccnagcGC aacggccATa tcggattggG
201  aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcgggccg
251  ttgaaggaaa TAtcgGctac attgtccgct tttccgatca cgggcacaaa
301  ttccattcgc ccttcGAcaa ccaTGCCTCA CATTCCGATT CTGACGAAGC
10  351  CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT
401  ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT
451  CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAGGCGGT
501  TGCCCAAAAT ATCCGCCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC
551  GGCTTGCCGA CCGTTTCCAC AATGCCGGCG CTATGCTGAC GCAAGGAGTA
601  GGCACGGAT TCAAAACGCG CACCCGATAC AGCCCCGAGC TGGACAGATC
15  651  GGGCAATGcc gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAAACA
701  TCATCGGCGC GGCAGGAGAA ATTGTcGGCG CAGGCGATGC CGTGCagGGT
751  ATAAGCGAAG GCTCAAACAT TGCTGTcATG CACGGCTTGG GTCTGCTTTC
801  CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC
851  TCAAAGACTA TGCCGcAGCA GCCATCCGCG ATTGGGCAGT CAAAAACCCC
20  901  AATGCCGCAC AAGGCATAGA AGCCGTcAGC AATATCTTTA TGGCAGCCAT
951  CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA
1001 TCACGGCACA TCCTGTCAAG CGGTcGCAGA TGGGCGCGAT CGCATTGCCG
1051 AAAGGGAAAT CCGCCGTcAG CGACAATTTT GCCGATGCGG CATAcGCCAA
1101 ATACCCGTCC CTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
25  1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
1201 AAAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
1301 AGTCGATAT TCAAGAATTA TCGGGGGCG GTATACCTAA GGCTAAGCCT
1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
30  1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA
1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAAATAAA
1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG
1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC
1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA
35  1651 ATTAAGAAAG CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA
1701 AGTGATGACC AAGCACACCA TGTTCCCAA AGATTGGGAT GAGGCTAGAA
1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
1801 AATAATGGC AGGTTACAAG TAAATCGGGT ATTAATAATAG AAGGATTTAC
40  1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 464; ORF46ng-1):

```

1  LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRQ HFEPDGKYHL
51  FGSRGELAXR NGHIGLGNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGHK
45  101  FHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
151  PAPKGARDIY SYDIKVAQN IRLNLTDNRS TGQRLADRFH NAGAMLTQGV
201  GDGFKRATRY SPELDRSGNA AEAFNGTADI VKNIIGAAGE IVGAGDAVQG
251  ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
301  NAAQGIeAVS NIFMAAIPK GIGAVRGKYG LGGITAHpVK RSQMGAIALP
351  KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
50  401  KNVKLADQRH PKTGVPFDGK GFPNFEKHVK YDTKLDIQEL SGGGIPKAKP
451  VFDAKPRWEV DRKLNKLtTR EQVEKNVQET RRRSQSSQFK AHAQREWENK
501  TGLDFNHFIG GDINKKGTVT GGHSLTRGDV RVIQQTsAPD KHGVYQATVE
551  IKKPDGSWEV KTKKGGKVMT KHTMFpKDWD EARIRAEVTS AWESRIMLKD
55  601  NKWQGTskSG IKIEGFTEPN RTAYPIYE*

```

ORF46ng-1 (SEQ ID NO: 464) and ORF46-1 (SEQ ID NO: 460) show 94.7% identity in 227 aa overlap:

5	orf46-1.pep		10	20	30	40
			AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER			
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFI	10	20	30	40
			RQVLDROHFEPDGKYHLFGSRGELAXR			
10	orf46-1.pep	50	60	70	80	90
		QSHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFS				
	orf46ng-1	50	60	70	80	90
		NGHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFS				
15	orf46-1.pep	110	120	130	140	150
		VDGFSLYRIHWDGYEHHHPADGYDGPQGGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS				
	orf46ng-1	110	120	130	140	150
		VDGFSLYRIHWDGYEHHHPADGYDGPQGGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS				
20	orf46-1.pep	170	180	190	200	210
		TGQRLADRFHNAGSMLTQGVGDGPKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE				
	orf46ng-1	170	180	190	200	210
		TGQRLADRFHNAGSMLTQGVGDGPKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE				
25	orf46-1.pep	I				
	orf46ng-1	I				
		250	260	270	280	290
		IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP				

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 (SEQ ID NO: 464) shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) (SEQ ID NO: 466) from strain A of *N. meningitidis*:

35	orf46a.pep	10	20	30	40	50
		LGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER				
	orf46ng-1	10	20	30	40	50
		LGISRKISLILSILAVCLPMHAHASDLANDPFI				
40	orf46a.pep	70	80	90	100	110
		SGHIGLGKIQSHQLGNLFIQQAIAKGNIGYIVRFS				
	orf46ng-1	70	80	90	100	110
		NGHIGLGKIQSHQLGNLFIQQAIAKGNIGYIVRFS				
	orf46a.pep	130	140	150	160	170
		VDGFSLYRIHWDGYEHHHPADGYDGPQGGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS				
	orf46ng-1	130	140	150	160	170
		VDGFSLYRIHWDGYEHHHPADGYDGPQGGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS				

-350-

		130	140	150	160	170	180
		190	200	210	220	230	240
	orf46a.pep	TGQRLVDRFHNTGSMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE					
5	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf46a.pep	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP					
10	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf46a.pep	NAAQGIEAVSNIFTAVIPVKIGIGAVRGKYGLGGITAHVPKRSQMGEIALPKGKSAVSDNF					
15	orf46ng-1	NAAQGIEAVSNIFMAAIPKIGIGAVRGKYGLGGITAHVPKRSQMGAIALPKGKSAVSDNF					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf46a.pep	ADAAYAKYPSPYHSRNIIRSLEQRYGKENITSSTVPPSNGKNVKLANKRHPKTKVPFDGK					
20	orf46ng-1	ADAAYAKYPSPYHSRNIIRSLEQRYGKENITSSTVPPSNGKNVKLADQRHPKTKVPFDGK					
		370	380	390	400	410	420
		430	440	450	460	470	
	orf46a.pep	GFPNFEKDVKYDTRINTAVPQVN---PIDEPVFN--PKGSVGSASHSWSITARIQYAKLP					
25	orf46ng-1	GFPNFEKHVKYDTKLD--IQELSGGGIPKAKPVFDAKPRWEVDKRLN-KLTREQVEKNV					
		430	440	450	460	470	
		480	490	500	510	520	530
	orf46a.pep	RQGRIRYIPPKNYSPSAPLPKGPNNGYLDKFGNEWTGKPSRTKGQEFEDVQLSKTGREQ					
30	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDVRIQQTS					
		480	490	500	510	520	530

The complete length ORF46a DNA sequence (SEQ ID NO: 465) is:

	1	TTGGGCATTT	CCCGCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
35	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGGTTCT	CGACCGTCAG	CATTTCTGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGAAC	TGCCGAGCGC	AGCGGTCATA	TCGGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
40	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCAGCAA
	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCCGCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGCGT
45	501	TGCCCCAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
	551	GGCTTGTCTG	CCGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCGGCG	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTATG	CACGCTTGG	GTCTGCTTTC
50	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC
	851	TCAAAGACTA	TGCCGCAGCA	GCCATCCGCG	ATTGGGCAGT	CCAAAACCCC
	901	AATGCCGCAC	AAGGCATAGA	AGCCGTCAGC	AATATCTTTA	CGGCAGTCAT

5  
10  
15

```

951 CCCCCTCAAA GGGATTGGAG CTGTTTCGGGG AAAATACGGC TTGGGCGGCA
1001 TCACGGCACA TCCTGTCAAG CGGTCGCAGA TGGGCGAGAT CGCATTGCCG
1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA
1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGA
1201 AAGAATGTGA AACTGGCAAA CAAACGCCAC CCGAAGACCA AAGTGCCGTT
1251 TGACGGTAAA GGGTTTCCGA ATTTTGAAAA AGACGTAAAA TACGATACGA
1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCTTT
1351 AATCCTAAAG GTTCTGTCGG ATCGGCTCAT TCTTGGTCTA TAACTGCCAG
1401 AATTCAATAC GCAAAATTAC CAAGGCAAGG TAGAATCAGA TATATCCAC
1451 CTAAAAATTA CTCTCCTTCA GCACCGCTAC CAAAAGGACC TAATAATGGA
1501 TATTTGGATA AATTTGGTAA TGAATGGACT AAAGGTCCAT CAAGAACTAA
1551 AGGTCAAGAA TTTGAATGGG ATGTTCAATT GTCTAAAACA GGAAGAGAGC
1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT
1651 GGAAAGATTA CACACAAATG A

```

This corresponds to the amino acid sequence (SEQ ID NO: 466):

20  
25

```

1  LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRO HFEPDGKYHL
51  FGSRGELAER SGHIGLGNIQ SHQLGNLFIQ QAAIKGNIGY IVRFSDHGHE
101 VHSFPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGGY
151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLVDRFH NTGSMLTQGV
201 GDGFKRATRY SPELDRSNA AEAFTGTADI VKNIIGAAGE IVGAGDAVQG
251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAQNP
301 NAAQIEAVS NIFTAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGEIALP
351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYKENI TSSTVPPSNG
401 KNVKLANKRH PKTKVPFDGK GFPNFEKDVK YDTRINTAVP QVNPIDEVPF
451 NPKGSVGSAA SWSITARIQY AKLPRQGRIR YIPPKNYSPS APLPKGPNNG
501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID
551 GKITHK*

```

30 Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 56

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 467):

35  
40

```

1  ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTT
51  GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTTTCGT
201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GGCGGTTTGT TTTGACGGGC
251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTG...

```

This corresponds to the amino acid sequence (SEQ ID NO: 468; ORF48):

45

```

1  MNIHTLLSKQ WTLPPFLPKR LLLSLLILLA PNAVFWVLAL LTATARPIVN
51  LDYLPALLI ALPWRFKIA GVLAFWLAVL FDGLMMVIQL PPFMDLIGAI
101 NLVPFILTAP APYQIMTGL...

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 469):

```

      1  ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTCT
      51  GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
5      101  TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
      151  TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCCTTTTCGT
      201  CAAAATTGCC GGCCTATTGG CGTTTGGCT GGCCTTTTG TTTGACGGGC
      251  TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
      301  AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
10     351  CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
      401  CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCGTTGTG
      451  GCGGCAGCGG GCTATTTTAC CGGCCATTTG AGTTACTACG ACCGGGGTCG
      501  GATGGCCAAAT ATCTTCGGCG CAAACAACCT CTACTACGCC AAAAGTCAGG
      551  CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
15     601  TGATGATCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
      651  TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
      701  GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
      751  CTGGCGCAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
      801  CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
20     851  GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
      901  TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
      951  CGGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001  GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
1051  GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTCTGGC AAGTGTCGGC
25     1101  ATTTTTCAAA AAACACGACA AGGGACTGTT TTACTGGATG ACGCTGACCA
      1151  GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
      1201  ACCGAATATG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTTACAGCT
      1251  GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
      1301  TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
30     1351  AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGCCTGGCT
1401  GAACTTCAAA ATCAAATAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 470; ORF48-1):

```

      1  MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFVVLAL LTATARPIVN
35     51  LDYLPALLI ALPWRFKIA GVLAFLAVL FDGLMMVIQL FPFMDLIGAI
      101  NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIIVCAAVV
      151  AAAGYFTGHL SYDRGRMAN IFGANFYA KSQAMLYTVS QNADFITAGL
      201  VDPVFLPLGN QORAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL
40     251  LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFAR
      301  CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKTC
      351  AIFGGVCDSE LFGEVSFFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC
      401  TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPPVG
      451  NLNETFRYLK QGHVAWLNFK IK*

```

45 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF48 (SEQ ID NO: 468) shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) (SEQ ID NO: 472) from strain A of *N. meningitidis*:

-353-

		10	20	30	40	50	60
	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFVWLALLTATARP	IVNLDYLP	PAALLI			
5	orf48a	MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVAVFWVLALLTATARP	IVNLXYLPAALLI				
		10	20	30	40	50	60
		70	80	90	100	110	119
	orf48.pep	ALPWRVFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAP	PAPYQIMTGL				
10	orf48a	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFIXTAP	ALYQIMTGLL				
		70	80	90	100	110	120
	orf48a	LLYMLAMPFVLQKAAAKTDFR	HIAACAAVVVAAGYFTGHL	SXYDRGRMANIFGANNFY	Y		
		130	140	150	160	170	180

The complete length ORF48a nucleotide sequence (SEQ ID NO: 471) is:

15	1	ATGAATATTC	ACACCCTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTCCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTNNCC	CCCAATGCGG
	101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCCGCC	GATTGTCAAT
	151	TTGGANTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTNTCGT
	201	CAAAATTGNC	GGCGTATTGG	CGTNTTGGCT	GGCGGTTTGT	TTTGACGGGC
20	251	TGATGATGGT	GATCCAACTC	TTCCCTTTTA	TGGATCTCAT	CGGCGCCATC
	301	AACCTCGTCC	CCTTCATCNT	GACCGCCCCC	GCCCTTTATC	AGATAATGAC
	351	CGGGCTGTTA	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAGAAAG
	401	CCGCCGCCAA	AACCGACTTC	CGACACATTG	CCGCCTGTGC	CGCCGTTGTG
	451	GTGGCAGCCG	GCTATTTTAC	CGGCCATTGT	AGTTANTACG	ACCGGGGGCG
25	501	GATGGCCAAT	ATCTTCGCGG	CAAACAACTT	CTATTACGCC	AAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGGCCTG
	601	GTCGATCCCG	TCTTCCTCCC	CTTGGGCAAT	CAACAGCGTG	CCGCCACGCA
	651	TCTGAACGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
	701	GGGGGCTGCC	GGCCAATCCC	GAACCTCAAA	ACGCCACTTT	TGCCAAACTG
30	751	CTGGCGCAAA	AAGANCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
	801	CATCGGCGCG	ACGATCGAAG	GCGAAATGCG	CGAACTGTGT	GCCTACGGCG
	851	GTTTGCGCGG	GTTTCGCACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
	951	CGGCGCGGGC	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
35	1001	GCTTTCAAGA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGGTGT	CGACAGCGAG	CTGTTTCGGC	AAGTGTCCGC
	1101	ANTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCNGACATTT	TCAACCACAG	GCTCAAATGC
	1201	ACCGAATATG	GCCTGCCCGC	CGAAACCGAC	NTCTGCCGCA	ATTTCAGCCT
40	1251	GCACACCCAA	TTCTTCGACC	AACTGGCGGA	TTTGATCCAA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCGTCGGC
	1351	AACCTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGGCACG	TCGNCTGGCT
	1401	GAACCTCAAA	ATCAATAA			

45 This encodes a protein having amino acid sequence (SEQ ID NO: 472):

	1	MNIHTLLSKQ	WTLPPFLPKR	LLLSLLILLX	PNAVFWVLAL	LTATARP	PIVN
	51	LXVLP	PAALLI	ALPWRXVKIX	<u>GVLAXWLAVL</u>	<u>FDGLMMVIQL</u>	<u>FPFMDLIGAI</u>
	101	NLVP	FIXTAP	ALYQIMTGLL	LLYMLAMPFV	LQKAAAKTDF	RHIAACAAVV
	151	VAAGYFTGHL	SXYDRGRMAN	IFGANNFY	Y	KSQAMLYTVS	QNADFITAGL
50	201	VDPVFLPLGN	QORAATHLNE	PKSQKILFIV	AESWGLPANP	ELQNATFAKL	
	251	LAQKXRFVSW	ESGSFPFIGA	TIEGEMRELC	AYGGLRGFAL	RRAPDEKFAR	
	301	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKTC	
	351	AIFGGVCDSE	LFGEVSAXFK	KHDKGLFYWM	TLTSHADYPE	SDIFNHRLKC	
	401	TEYGLPAETD	XCRNFSLHTQ	FFDQLADLIQ	RPEMKGTEVI	IVGDHPPVVG	

-354-

451 NLNETFRYLK QGHVXWLNFK IK\*

ORF48a (SEQ ID NO: 472) and ORF48-1 (SEQ ID NO: 470) show 96.8% identity in 472 aa overlap:

```
5.      10      20      30      40      50      60
orf48a.pep MNIHTLLSKQWTLPPFLPKRLLSLLILLXPNAVFWVLALLTATARPVNLXYLPAALLI
orf48-1     MNIHTLLSKQWTLPPFLPKRLLSLLILLAPNAVFWVLALLTATARPVNLXYLPAALLI
           10      20      30      40      50      60

10      70      80      90     100     110     120
orf48a.pep ALPWRXVKIXGVLAXWLAFLDGLMMVIQLFPFMDLIGAINLVPFXTAPALYQIMTGLL
orf48-1     ALPWRXVKIXGVLAXWLAFLDGLMMVIQLFPFMDLIGAINLVPFXTAPALYQIMTGLL
           70      80      90     100     110     120

15      130     140     150     160     170     180
orf48a.pep LLYMLAMPFVLQKAAAKTDFRHHIAACAAVVVAAGYFTGHLSXYDRGRMANIFGANNFYA
orf48-1     LLYMLAMPFVLQKAAAKTDFRHHIAACAAVVVAAGYFTGHLSXYDRGRMANIFGANNFYA
           130     140     150     160     170     180

20      190     200     210     220     230     240
orf48a.pep KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNEPKSQKILFIVAESWGLPANP
orf48-1     KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRATHLNEPKSQKILFIVAESWGLPANP
           190     200     210     220     230     240

25      250     260     270     280     290     300
orf48a.pep ELQNATFAKLLAQKXRFVSWESGSFPFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA
orf48-1     ELQNATFAKLLAQKXRFVSWESGSFPFIGATIEGEMRELCAYGGLRGFALRRAPDEKFA
           250     260     270     280     290     300

30      310     320     330     340     350     360
orf48a.pep CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAENLIGKKTCAIFGGVCDSE
orf48-1     CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAENLIGKKTCAIFGGVCDSE
           310     320     330     340     350     360

35      370     380     390     400     410     420
orf48a.pep LFGEVSAXFKKHDKGLFYWMTLTSHADYPESDIFNHRLKCTEYGLPAETDXCRNFSLHTQ
orf48-1     LFGEVSAXFKKHDKGLFYWMTLTSHADYPESDIFNHRLKCTEYGLPAETDXCRNFSLHTQ
           370     380     390     400     410     420

40      430     440     450     460     470
orf48a.pep FFDQLADLIQRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX
orf48-1     FFDQLADLIQRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX
           430     440     450     460     470
```

45 Homology with a predicted ORF from *N.gonorrhoeae*



ORF48 (SEQ ID NO: 468) shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) (SEQ ID NO: 474) from *N. gonorrhoeae*:

orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL DYLP AALLI	60
5	:    :	
orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNL DYLP AALLI	60
orf48.pep	ALPWRVFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
orf48ng	ALPWRVFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL	120

10 The ORF48ng nucleotide sequence (SEQ ID NO: 473) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 474):

1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVFWVLAL	LTATARPIVN
51	LDYLP AALLI	ALPWRVFKIA	GVLAFWPAVL	FDGLMMVIQL	FPFMDLIGAI
101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQKAAVKTDF	RHIAVCAAVV
15	151	AAARYFTGPF	ELLRTGGRWQ	YVQHRRLLS	GSRASFRRRQ
	201	PYASMGNGG..			KADVLRLRGN

Further work identified the complete gonococcal DNA sequence (SEQ ID NO: 475):

20	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTTCCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
	101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCCGCC	GATTGTCAAT
	151	TTGGACTACC	TCCCGCCGCG	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
	201	CAAAATTGCC	GGCGTATTGG	CGTTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAACCT	TCCCTTTTA	TGGACCTCAT	CGGCGCCATC
25	301	AACCTCGTCC	CCTTCATCCT	GACCGCCCCC	GCCCCTTATC	AGATAATGAC
	351	CGGGCTGTTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
	401	CCGCCGTCAA	AACCGACTTC	CGACACATTG	CCGTCTGTGC	CGCCGTTGTG
	451	GCGGCAGCCG	GCTATTTTAC	CGGCCATTTC	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAT	ATCTTCGGCG	CAAAACAATT	CTATTACGCC	aaAAGTCAGG
30	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgectG
	601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CCGCCACGCG
	651	GCTGAGTGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
35	801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTGCGCGG	GTTCGCACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
	951	CGGCGCGGGT	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
40	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTCCGGC	AAGTGTCCGC
	1101	ATTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
	1201	ACCGAATACG	GCCTGCCCGC	CGAAACCGAC	CTCTGCCGCA	ATTTCAGCCT
	1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
45	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCCCGTCGC
	1351	AACTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGACACG	TGCCTGGCT
	1401	GCACTTCAAA	ATCAATAA			

This encodes a protein having amino acid sequence (SEQ ID NO: 476; ORF48ng-1):

5  
10  
15  
20  
25  
30  
35  
40  
45

1 MNIHALLSEQ WTLPPFLPKR LLLSLILLAPNAVFWVLAL LTATARPIVN  
51 LDYLPALLI ALPWRFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI  
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD RHIACAAVV  
151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL  
201 VDPVFLPLGN QORAATRLSE PKSQKILFIV AESWGLPGNP ELQNATFAKL  
251 LAQKDRFSVW ESGSFPIGA TVEGEMRELC AYGGRLGFAL RRAPDEKFA  
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQIKT AENLIGKKT  
351 AIFGGVCDSE LFGEVSAFFK KHDGKLFYWM TLTSHADYPE SDIFNHLK  
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPV  
451 NLNETFRYLK QGHVAWLHFK IK\*

ORG48ng-1 (SEQ ID NO: 476) and ORF48-1 (SEQ ID NO: 470) show 97.9% identity in 472 aa overlap:

15  
20  
25  
30  
35  
40  
45  
50

orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1  
orf48-1.pep  
orf48ng-1

10 20 30 40 50 60  
10 20 30 40 50 60  
70 80 90 100 110 120  
70 80 90 100 110 120  
130 140 150 160 170 180  
130 140 150 160 170 180  
190 200 210 220 230 240  
190 200 210 220 230 240  
250 260 270 280 290 300  
250 260 270 280 290 300  
310 320 330 340 350 360  
310 320 330 340 350 360  
370 380 390 400 410 420  
370 380 390 400 410 420  
430 440 450 460 470  
FFDQLADLIQRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVAWLNFKIKX

orf48ng-1 FFDQLADLIRRP<sup>430</sup>EMKGTEVI<sup>440</sup>IIVGDHPPVPGN<sup>450</sup>LN<sup>460</sup>ETFRYLKQGHV<sup>470</sup>AWLHFKIKX

5 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 57

10 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 477):

15

```

      1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCCAAAA TCATCATCGT
     51  TACTTTGAGT ATCGCCACGC TTGCCGCCGC CGGCATCGCT ATGTCGCGCG
    101  GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
    151  GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
    201  AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
    251  CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
    301  AGTGCGGTTT TGGCTTTGGT TTTCCTTGCA CTGGGCGC.G TAGCGCCGAA
    351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAAATAT AACGGGCAAT
    401  TGATCAATAT GTACGCC..

```

20

This corresponds to the amino acid sequence (SEQ ID NO: 478; ORF53):

25

```
      1  ..VSGGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMQSD FIEPTPWTLA
     51  GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
    101  SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NGQLINMYA..
```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 479):

	1	ATGTCCGAAC	AACATATTTTC	GACTTGGAAA	AGTAAAATCA	ACGCATTGGG
	51	TCCGGGGATC	ATGATGGCTT	CGGCGGCGGT	CGGCGGTTTC	CACCTGATTG
30	101	CCTCGACGCA	GGCGGGCGCG	CTTTACGGCT	GGCAGATCGC	GCTCATCATC
	151	ATCCTGACCA	ACCTCTTCAA	ATACCCGTTT	TTCGCTTCA	GCGCGCATTA
	201	CACGCTGGAC	ACGGGCAAGA	GCCTGATTGA	AGGTTATTGC	GAGAAAAGCC
	251	GCCTTTTATT	TGGGTATTTC	CTGATTTTGT	GACTCTCTC	CGCCACGATT
	301	AACGCGGGCG	CGGTGCGCAT	TGTAACCGCC	GCCATCGTCA	AAATGGCGAT
35	351	TCCCTCGCTG	ATGTTTGATG	CCGGCACGGT	TGCCGCCTTG	ATTATGGCAT
	401	CCTGCCTGAT	TATTTTGGTG	AGCGGACGTT	ACCGCGCTTT	GGATCGCGTT
	451	TCCAAAATCA	TCATCGTTAC	TTTGAGTATC	GCCACGCTTG	CCGCCGCCGG
	501	CATCGCTATG	TCGCGCGGTA	TGCAGATGCA	GTCCGATTTT	ATCGAGCCGA
	551	CACCGTGGAC	GCTTGCCGGT	TTGGGCTTCC	TGATCGCGTT	GATGGCGCTG
40	601	ATGCCCGCGC	CGATTGAAAT	TTCCGCCATC	AATTTCTTGT	GGGTAACCGA
	651	AAAACAACGC	ATCAATCCTT	CCGAATACCG	CGACGGGATT	TTTGATTTCA
	701	ACGTCGGTTA	TATCGCCAGT	GCGGTTTTTG	CTTTGGTTTT	CCTTGCACTG
	751	GGCGCGTTTG	TGCAATACGG	CAACGCGGAA	GCAGTGCAGA	TGGCGGGCGG
	801	CAAATATATC	GGGCAATTGA	TCAATATGTA	CGCCGTACC	ATCGGCGGCT
45	851	GGTCGCGCCC	GCTGGTGGCG	TTTATCGCGT	TTGCCTGTAT	GTACGGCACG
	901	ACGATTACCG	TCGTGGACGG	CTATGCCCGT	GCCATTGCCG	AACCCGTGCG
	951	CTGTCTGCGC	GGAAAAGACA	AAACGGGCAA	CGCCGAATTC	TTTGCTTGGA
	1001	ATATTTGGGT	GGCGGGCAGC	GGTTTGGCGG	TGATTTTCTG	GTTTGACGGC

-358-

1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC  
 1101 CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTAAAGGT GATGAAAAAC  
 1151 ACAAACTCAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTTAT  
 1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA  
 1251 ATGA

This corresponds to the amino acid sequence (SEQ ID NO: 480; ORF53-1):

1 MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASQAGA LYGWQIALII  
 51 ILTNLFKYPP FRFSAHYTLD TGKSLIEGYA EKSRVYLWVF LILCILSATI  
 101 NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV  
 151 SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW  
 201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL  
 251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT  
 301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIWVAGS GLAVIFWFDG  
 351 VMANLLKFAM IAAFVSAPVF AWLNRYLVKG DEKHKLTSGM NALALAGLIY  
 401 LTGFTVLFL NLAGMFK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF53 (SEQ ID NO: 478) shows 93.5% identity over a 139aa overlap with an ORF (ORF53a)

(SEQ ID NO: 482) from strain A of *N. meningitidis*:

					10	20	30
orf53.pep					VSGRYRALDRVSKIIIVTSLIATLAAAGIA		
orf53a	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVS				GRYRALDRVSKIIIVTSLIATLAAAGIA		
	110	120	130	140	150	160	
		40	50	60	70	80	90
orf53.pep		MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG					
orf53a	MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG						
	170	180	190	200	210	220	
		100	110	120	130	139	
orf53.pep		IFEFNVGYIASAVLALVFLALGXVAPNGXGTVQMAGGKYNGQLINMYA					
		:					
orf53a	IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLV						
	230	240	250	260	270	280	
orf53a	AFIAFACMYGTTITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFD						
	290	300	310	320	330	340	

The complete length ORF53a nucleotide sequence (SEQ ID NO: 481) is:

1 ATGTCCGAAC AACATATTTT GACTTGAAA AGTAAATCA ACGCATTGGG  
 51 ACCGGGGATT ATGATGGCTT CGGCGGCGGT CGGCGGTTCG CACCTGATTG  
 101 CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC  
 151 ATCCTGACCA ACCTCTCAA ATACCGGTTT TTCCGCTTCA GCGCGCATT  
 201 CACGCTGGAC ACGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC  
 251 GCGTTTATT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT  
 301 AACGCGGGCG CGGTCGCCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT

5  
10  
15  
20

```

351 TCCCTCGCTG ATGTTTGTATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
401 CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
451 TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601 ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTC
701 ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CCTTGCAC
751 GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
851 GGTCGCGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCACG
901 ACGATTACCG TTGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG
951 CCTGCTGCGC GGAAAGACA AAACGGGCAA CGCCGAATTC TTTGCCTGGA
1001 ATATTTGGGT GGCGGGCAGC GGTTTGGCGG TGATTTTCTG GTTTGACGGC
1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101 CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTCAAAGGT GATGAAAAAC
1151 ACAAACCTAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTTAT
1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251 ATGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 482):

25  
30

```

1  MSEQHISTWK SKINALGPGI MMASAAVGS HLIASTQAGA LYGWQIALII
51  ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSRYVLWVF LILCILSATI
101 NAGAVAIUTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
151 SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL
251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIIWVAGS GLAVIFWFDG
351 VMANLLKFAM IAAFVSAPVF AWLNRYRLVKG DEKHKLTSKM NALALAGLIY
401 LTGFTVLFLN NLAGMFK*

```

ORF 53a (SEQ ID NO: 482) shows 100.0% identity in 417 aa overlap with ORF53-1 (SEQ ID NO: 480):

35  
40  
45  
50

```

              10      20      30      40      50      60
orf53a.pep  MSEQHISTWKS KINALGPGI MMASAAVGS HLIASTQAG ALYGWQIALII ILTNLFKYPF
              |||||||
orf53-1     MSEQHISTWKS KINALGPGI MMASAAVGS HLIASTQAG ALYGWQIALII ILTNLFKYPF
              10      20      30      40      50      60

              70      80      90     100     110     120
orf53a.pep  FRFSAHYTLD TGKSLIEGYA EKSRYVLWV FLILCILSATI NAGAVAIUTA AIVKMAIPSL
              |||||||
orf53-1     FRFSAHYTLD TGKSLIEGYA EKSRYVLWV FLILCILSATI NAGAVAIUTA AIVKMAIPSL
              70      80      90     100     110     120

              130     140     150     160     170     180
orf53a.pep  MFDAGTVAAL IMASCLIIIV SGRYRALDRV SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF
              |||||||
orf53-1     MFDAGTVAAL IMASCLIIIV SGRYRALDRV SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF
              130     140     150     160     170     180

              190     200     210     220     230     240
orf53a.pep  IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEK QRINPSEYRD GI FDFNVGYIAS
              |||||||
orf53-1     IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEK QRINPSEYRD GI FDFNVGYIAS

```

-360-

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf53a.pep	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT					
	orf53-1	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT					
		250	260	270	280	290	300
		310	320	330	340	350	360
10	orf53a.pep	TITVVDGYARAIAPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM					
	orf53-1	TITVVDGYARAIAPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM					
		310	320	330	340	350	360
		370	380	390	400	410	
15	orf53a.pep	IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMPFKX					
	orf53-1	IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMPFKX					
		370	380	390	400	410	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 (SEQ ID NO: 478) shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) (SEQ ID NO: 484) from *N. gonorrhoeae*:

20	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVS	91
	orf53.pep	MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
25	orf53ng	MSRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA	139
	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYAVTIGGWSRPLV	211

An ORF53ng nucleotide sequence (SEQ ID NO: 483) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 484):

	1	MPKKSCVYLW	VFLILCIASA	TINAGAVAIV	TAAIVKMAIP	SLMFDAGTVA
	51	ALIMASCLII	LVSGRYRALD	RVSKIIIVTL	SIATLAAAGI	AMSRGMQMQP
	101	DFIEPTPWT	AGLGFLIALM	GWMPAPIEIS	AINSLWVTEK	QRINPSEYRD
	151	GIFDFNVGYI	ASAVLALVFL	ALGAFVQYGN	GEAVQMGGGK	YIGQLINMYA
35	201	VTIGGWSRPL	VAFIAFACMY	GAASTVVDGY	ARAIAPVRL	LRGKDKTARP
	251	IVLLEKLGR	HRFGRDFLV*			

Further analysis revealed further partial DNA gonococcal sequence (SEQ ID NO: 485):

	1	..aagaAAAGCT	GCGTTTATTT	GTGGGTTTTT	TTGATTTTGT	GTATCGCCTC
40	51	CGCCACGATT	AACGCGGGCG	CGTCGCCAT	TGTAACCGCC	GCCATCGTCA
	101	AAATGGCGAT	TCCCTCGCTG	ATGTTTGATG	CCGGCACGGT	TGCCGCCTTG

5  
10  
15

```

151  ATTATGGCAT CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT
201  GGATCGTGTT TCCAAAATCA TCATTGTTAC TTTGAGCATC GCCACGCTTG
251  CCGCCGCCGG CATCGCTATG TCGCGCGGTA TGCAGATGCA GCCCGATTTT
301  ATCGAGCCGA CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT
351  GATGGGCTGG ATGCCCGCGC CGATCGAAAT TTCCGCCATC AATTCTTTGT
401  GGGTAACCGA AAAACAACGC ATCAATCCTT CTGAATACCG CGACGGGATT
451  TTCGATTTCa ACGTCGGTTA TATCGCcagT GCGGTTTTTG CTTTGGTTTTT
501  CCTTGCACTG GCGCGCTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA
551  TGGCGGGCGG CAAATATATC GGGCAATTGA TTAATATGTA TGCCGTAACC
601  ATCGGCGGCT GGTCTCGTCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT
651  GTACGGCAGC ACGATTACCG TTGTGGACGG TTATGCGCGT GCCATTGCCG
701  AACCCGTGCG CCTGTGCGC GGCAGGGATA AAACGGCAA CGCCGAGTTG
751  TTTgccTGGa ATATTGGGGT GGCGGGCAGC GGTTTGGCGG TGATTTTCTG
801  GTTTGACggc gcaaTGGCgG AAcTgcTCAa ATTTGCGATG AtTgccgcCT
851  TTGTGTCCGC CCCTGTGTTT GCCTGGCTCA ACTACCGCCT CGTCAAAGGG
901  GACAAACGCC ACAGGCTTAC CGCCGGTATG AACGCCCTTG CCATTGTCTG
951  CCTGCTCTAC CTGGCCGGGT TTGCCGTTTT GTTCTGTGTT AACCTTACCG
1001 GACTTTTGGC ATAG

```

20 This corresponds to the amino acid sequence (SEQ ID NO: 486; ORF53ng-1):

25

```

1  ..KKSCVYLWVF LILCIASATI NAGAVAIvTA AIVKMAIPSL MFDAGTVAAL
51  IMASCLIIlV SGRYRALDRV SKIIIVTLsI ATlAAAGIAM SRGMQMqPDF
101 IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEKQR INPSEYRDGI
151 FDFNVGYIAS AVLALVFLAL GAFVQYNGE AVQMAGGKYI GQLINMYAVT
201 IGGWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL
251 FAWNIIWAGS GLAVIFWFDG AMAELLKFAM IAAFVSAPVF AWLNYRLVKG
301 DKRHRLTAGM NALAIVGLLY LAGFAVLFLl NLTGLLA*

```

30 ORF53ng-1 (SEQ ID NO: 486) and ORF53-1 (SEQ ID NO: 480) show 94.0% identity in 336 aa overlap:

35

```

           60      70      80      90      100     110
orf53-1.pep  ILTNLFKYPPFRFSaHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIvTA
orf53ng-1      :||| ||||| ||||| ||||| ||||| |||||
                10      20      30

           120     130     140     150     160     170
orf53-1.pep  AIVKMAIPSLMFDAGTVAALIMASCLIIlVSGRYRALDRVSKIIIVTLsIATlAAAGIAM
orf53ng-1      ||||| ||||| ||||| ||||| ||||| |||||
                40      50      60      70      80      90

           180     190     200     210     220     230
orf53-1.pep  SRGMQMqSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI
orf53ng-1      ||||| ||||| ||||| ||||| ||||| |||||
                100     110     120     130     140     150

           240     250     260     270     280     290
orf53-1.pep  FDFNVGYIASAVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVA
orf53ng-1      ||||| ||||| ||||| ||||| ||||| |||||
                160     170     180     190     200     210

```

50

		300	310	320	330	340	350
	orf53-1.pep	FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKDTGNAEFFAWNIWVAGSGLAVIFWFDG					
5	orf53ng-1	FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKDTGNAELFAWNIWVAGSGLAVIFWFDG					
		220	230	240	250	260	270
		360	370	380	390	400	410
	orf53-1.pep	VMANLLKFAMIAAFVSAPVFAWLNYRLVKGD <del>EXHKL</del> TSGMNALALAGLIYLTGFTVLFL					
10	orf53ng-1	AMAELLKFAMIAAFVSAPVFAWLNYRLVKGD <del>KRHL</del> TAGMNALAI <del>VGL</del> LYLAGFAVLFL					
		280	290	300	310	320	330
	orf53-1.pep	NLAGMPKX					
		: ::					
	orf53ng-1	NLTGLLAX					

- 15 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 58

- 20 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 487):

1 ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG  
51 TGCGCTTGCC GGCTTGTTTT TGTCCGCGC ACAATCCGAA CGCGAGTGGA  
101 TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG  
25 151 CTGCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT  
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTCTG  
251 TCCGTTTCTG CCGAAACTAT CTGGCGCACG AATCCGAACC GGACAGGCCC  
301 GTTCCGCCT..

This corresponds to the amino acid sequence (SEQ ID NO: 488; ORF58):

30 1 ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QEKKGEKQAE  
51 LPEIKDGMPP FPELALMLFH AVKTAVYWLF VGVRVFCRNY LAHESEPDPR  
101 VPP..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 489):

35 1 ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT  
51 GTTTTTTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG  
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA  
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT  
201 CAAAACGGCA GTGTATTGGC TGTTTGTCTG TGTCGTCCGT TTCTGCCGAA  
40 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT  
301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG  
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG  
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCCG



5 451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA  
 501 AATTTTCGCCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA  
 551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC  
 601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCGGA  
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC  
 701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG  
 751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC  
 801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC  
 851 ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC  
 10 901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCGCCG  
 951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG  
 1001 CGCGAATTC TCGCTGATT CCGGAAAGTC AGACGGTTGT CGGGAAACGG  
 1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAA GTTTTCACGG AAACCGTTTC  
 1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAACTGCC GATATCCATA  
 15 1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG  
 1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC  
 1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTC GAGCAGGTGC  
 1301 AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTTGAAT  
 1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG  
 20 1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCCGAAACC GAAGCGTTCCG  
 1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC  
 1501 CCGTCCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCCATC  
 1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC  
 1601 TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG  
 25 1651 GAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT  
 1701 CAAGGTTGTC GATTCTTATT CCGGCCCGGT AATTACGCGT TATGAAATCG  
 1751 AACCCGATGT CGGCGTGCGC GGCAATTCGG TTCTGAATCT GGAAAAAGAT  
 1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCCC  
 1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA  
 30 1901 TACGCTGAG CGAAATCTC AATTGCGCCG AGTTTGCCGA ATCCAAATCC  
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC  
 2001 CGACTTGGA AAAGCACCGC ATTTGTTGGT TGCCGGCACG ACCGGTCCGG  
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC  
 2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT  
 35 2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTCT GTTACCGATA  
 2201 TGAAGCTGGC GGCAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA  
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CGGGCTTCAA  
 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT  
 2351 TCAGCCTCAC GCCGACGAT CCCGAACCTT TGGAAAACT GCCGTTTATC  
 40 2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA  
 2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA  
 2501 TCCATTTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT  
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA  
 2601 AATCGACAGC CGCACGATTC TCGACCAAAT GGGCGCGGAA AACCTGCTCG  
 45 2651 GTCAGGGCGA TATGCTGTTC CTGCTGCCGG GTACTGCCTA TCCGACGCGC  
 2701 GTTACGCGC CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCAATA  
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG  
 2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGACGAAACC  
 2851 GATCCGATGT ACGCAGAGC CGTATCCGTT GTCCTGAAA CGCGCAAAGC  
 50 2901 CAGATTTTCG GCGGTACAGC GCGCCTTGC TATCGGCTAC AACCGCGCCG  
 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA  
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence (SEQ ID NO: 490; ORF58-1):

55 1 MFWIVLVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPIK  
 51 DGMPDFPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS  
 101 ANRADVPTAS DGYSDSNGNT EEAETEEAEA AEEEEADTED IATAVIDNRR  
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI  
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE

-364-

5 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQGKG QAEAKSPDVS  
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR  
 351 DVEMPSETEN VFTETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPPEV  
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPPSGF EQVQSRIAE TDHLADDVLN  
 451 GGWQEETA AI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFENVPSER  
 501 PSCRVSDETA DEGAFPSSET GAVSEHLPTT DLLLPPLFNP EATQTEEEEL  
 551 ENSITIEEKL AEFKVKVKV DSYS GPVITR YEIEPDVGVR GNSVLNLEKD  
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS  
 10 651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TSGGKSVGVN AMILSMLFKA  
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMKLAANA LNWCVNEMEK  
 751 RYRLMSFMGV RNLAGFNQKI AEAAARGEKI GNPFSLT PDD PEPLEKLPFI  
 801 VVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG  
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LLPGTAYPQR  
 901 VHGAFADEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDDET  
 15 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE  
 1001 HNGNRTILVP LDNA\*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

## 20 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 (SEQ ID NO: 488) shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) (SEQ ID NO: 492) from strain A of *N. meningitidis*:

25 orf58.pep LRETAYVLDSFD RYFVVALAGLFFVRAQSEREW MREVS AWQEKKG EKQAELPEIKDGM PD  
 orf58a MFWIVLIVILL LALAGLFFVRAQSEREW MREVS AWQEKKG EKQAELPEIKDGM PD  
 30 orf58.pep FPELALMLFH AVKTAVYWL FVG VVRFCR NYLAHESE PDRPVPP  
 orf58a FPELALMLFH AVKTAVYWL FVG VVRFCR NYLAHESE PDRPVPPASANRADVPTASDGYSD

The complete length ORF58a nucleotide sequence (SEQ ID NO: 491) is:

35 1 ATGTTTGGG TAGTTTGGT CGTTATTTG TTGCTTGCGC TTGCCGGCTT  
 51 GTTTTTGTC CGCGACAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG  
 101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA  
 151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT  
 201 CAAAACGGCA GTGTATTGGC TGTGTGTCGG TGTCGTCCGT TTCTGCCGAA  
 40 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT  
 301 GCAAATCGTG CCGATGTTCC GACCGCATCC GACGATATT CAGACAGTGG  
 351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG  
 401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC  
 45 451 ATCCCATTCG ACCGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA  
 501 AATTTGCCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA  
 551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC  
 601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA  
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC  
 701 AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAAGAAGC GTTTTCCGAG

5  
10  
15  
20  
25  
30  
35  
40  
45

```

751 TCTGCGGATT AC GGATT TGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC
851 ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
1001 CGCGGATTTT GCGCCTGATT CCGGAAAGTC GGACGGTTGT CCGGAAACGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAANTGTTTC
1101 GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
1201 CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCCGC CTCCCGTATC
1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTG GAGCAGGTGC
1301 AACCGAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGAGGG
1401 TGTGCGAGAG CGGTCAAGCG GGCAATATTT GTCGGAACCG GAAGCGTTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCGGTTTG AAAATGTGCC GTCTGAACCG
1501 CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCGCCGCT GTTCAATCCC GGGGCGACGC AAACCGAAGA AGANCTGTTG
1651 GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTGTGC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG
1751 AACCCGATGT CGGCGTGCGC GGCAATTCGG TTCTAAATCT GGAAAAAGAN
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCT AAACCATCCT
1851 CGGCAAAACC TGTATGGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA
1901 TACGCTGAG CGAAATCTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
2001 CCACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCACG ACCGGTTCGG
2051 GCAAAATCGGT GGGTGTC AAC GCGATGATTG TGTCTATGCT TTCAAAGCC
2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCGTGC GTTACCGATA
2201 TGAAGCTGGC GGCAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA
2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CCGGTNTCAA
2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAATC GGCAACCCGT
2351 TCAGCTCAC GCCGACAAT CCCGAACCTT TGGANAAATT GCCGTTTATC
2401 GTGGTCGTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501 TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCACGGGT
2551 CTGATTAAAG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCAGCAA
2601 AATCGACAGC CGCACGATTC TTGACCAAAT GGGTGCGGAA AACCTGCTCG
2651 GGCAGGGCGA TATGCTGTTT CTGCCGCCGG GTACGGCCTA TCCGCAGCGC
2701 GTTCAGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
2751 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2801 GTATGTCCGA CGATTGCTG GGAATCAGCC GGAGCGGCGA CGGCGAAACC
2851 GATCCGATGT ACGACGAGGC CGTGTCTGTT GTTTTGAAAA CGCGCAAAGC
2901 CAGCATTTCT GCGGTGCAGC GCGCATTGCG TATCGGCTAT AATCGCGCCG
2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACC GGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 492):

50  
55

```

1 MFWIVLIVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAE LPEIK
51 DGMPDPPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS
101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEAADTED IATAVIDNRR
151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE
251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQGKG QAEAKSPDVS
301 QGQSVSDGTA VRDAXRRVSV NLKEPNKATV SAEARISR LI PESRTVVGKR
351 DVEMPS ETEN VFTEXVSSVG YGXPVYDETA DIHIEEPAAP WDAWVVEP EV
401 PKVPM PAXDI PPPPVSEIY NRTYEPPAGF EQVQSR IAE TDHLADDVLN
451 GGWQEETA AI ANDGSEGV AE RSSGQYLSET EAFGHDSQAV CPFENVPSER
501 PSRRAXDTEA DEGA FQSEET GAVSEHLPTT DLLLPPLFNP GATQTEEXLL

```

5 551 XNSITIEEKX AEFKVKVKVV DSYSGPVITR YEIEPDVGVR GNSVLNLEKX  
601 LARSLGVASI RVVETILGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS  
651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TGSGKSVGVN AMILSMLFKA  
701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMKLAANA LNWCVNEMEK  
751 RYRLMSFMGV RNLAGXNQKI AEAAARGEKI GNPFSLTPDN PEPLXKL~~PFI~~  
801 ~~VVV~~DEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG  
851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LPPGTAYPQR  
901 VHGAFADEE VHRVVEYLKQ FGEPTYVDDX LSGGMSDDL GISRSGDGET  
951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE  
10 1001 HNGNRTILVP XDNA\*

ORF58a (SEQ ID NO: 492) and ORF58-1 (SEQ ID NO: 490) show 96.6% identity in 1014 aa overlap:

15 orf58a.pep 10 20 30 40 50 60  
MFWIVLIVILLALAGLFFVRAQSEREWREVS~~AWQ~~EKKGEKQAE~~LPEIKD~~GMPDFPELA  
|||  
orf58-1 MFWIVLIVILLALAGLFFVRAQSEREWREVS~~AWQ~~EKKGEKQAE~~LPEIKD~~GMPDFPELA  
10 20 30 40 50 60

20 orf58a.pep 70 80 90 100 110 120  
LMLFHAVKTAVYWLFGVVRFCRNYLAHESE~~PD~~RPVPPASANRADVPTASDGYSDSGNGT  
|||  
orf58-1 LMLFHAVKTAVYWLFGVVRFCRNYLAHESE~~PD~~RPVPPASANRADVPTASDGYSDSGNGT  
70 80 90 100 110 120

25 orf58a.pep 130 140 150 160 170 180  
EEAETEEAEEAEEAADTEDIATAVIDNRRIP~~FDR~~SIAEGLMPSESEIS~~PVR~~PVFKEITL  
|||  
orf58-1 EEAETEEAEEAEEAADTEDIATAVIDNRRIP~~FDR~~SIAEGLMPSESEIS~~PVR~~PVFKEITL  
130 140 150 160 170 180

30 orf58a.pep 190 200 210 220 230 240  
EEATRALNSAALRETKKRYIDAFEKNETAV~~PKVR~~VS~~DT~~PM~~EGLQ~~II~~GLDD~~PVLQRTYSRM  
|||  
orf58-1 EEATRALNSAALRETKKRYIDAFEKNETAV~~PKVR~~VS~~DT~~PM~~EGLQ~~II~~GLDD~~PVLQRTYSHM  
190 200 210 220 230 240

35 orf58a.pep 250 260 270 280 290 300  
FDADKEAFSESADYGFEPYFEKQHPSAFSA~~VKA~~ENARNAPFRRHAGQGKGQAEAKSPDVS  
|||  
orf58-1 FDADKEAFSESADYGFEPYFEKQHPSAFSA~~VKA~~ENARNAPFRRHAGQGKGQAEAKSPDVS  
250 260 270 280 290 300

40 orf58a.pep 310 320 330 340 350 360  
QQQSVSDGTAVRDAXRRVSVNLKEPNKATV~~SAE~~ARISRLIPESRTVVGKRDVEMPSETEN  
|||  
orf58-1 QQQSVSDGTAVRDARRRVSVNLKEPNKATV~~SAE~~ARISRLIPESQTVVGKRDVEMPSETEN  
310 320 330 340 350 360

45 orf58a.pep 370 380 390 400 410 420  
VFTEXVSSVGYGXPVYDETADIIHEEPAAP~~DAW~~VVEPPEVPKVPMPAXDIPPPPPVSEIY  
|||  
orf58-1 VFTETVSSVGYGXPVYDETADIIHEEPAAP~~DAW~~VVEPPEVPKVPMTAIDIQPPPPVSEIY  
370 380 390 400 410 420

50 orf58a.pep 430 440 450 460 470 480  
NRTYEPPAGFEQVQRSRIAETDHLADDVLN~~GGW~~QEETA~~AI~~ANDGSEGVAERS~~SS~~QYLSET

[illegible]

Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 (SEQ ID NO: 488) shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) (SEQ ID NO: 494) from *N. gonorrhoeae*:

```

      orf58.pep      ALMLFHAVKTAVYWLFVGVVRFRCRNYLAHESEPDRPVPP      103
                    |||||
5      orf58ng      SEPDRPVPPASANRADVPTASDGYSDSGNG      30

```

The ORF58ng nucleotide sequence (SEQ ID NO: 493) is predicted to encode a protein having partial amino acid sequence (SEQ ID NO: 494):

```

10      1  ..SEPDRPVPPA SANRADVPTA SDGYSDSGNG TEEAETEAAE AAEAAAADTE
      51  DIATAVIDNR RIPFDRSIAE GLMQSESKTS PVRPVFKEIT LEEATRALSS
     101  AALRETKKRY IDAFEKNGTA VPKVRVSDTP MEGLQIIGLD DPVLQRTYSR
     151  MFDADKEAFS ESADYGFEFY FEKQHPSAFS AVKAENARNA PFRRHAGQEK
     201  GQAEAKSPDV SQGQSVSDGT AVRDAARRVS VNLKEPNKAT VSAEARISRL
     251  IPESRTVVGK RDVEMPSETE NVFTETVSSV GYGGPVYDEA ADIHIEEPAA
     15  301  PDAWVVEPPE VPEVAVPEID ILPPPPVSEI YNRTYEPPAG FEQAQRSRIA
     351  ETDHLAADVL NGGWQEETAA IADDGSEGAA ERSSGQYLSE TEAFGHDSQA
     401  VCPFEDVPSE RPSCRVSDTE ADEGAFQSEE TGAVSEHLPT TDLLLPLPFN
     451  PEATQTEEL LENSITIEEK LAEFKVKVKV VDSYSGPVIT RYEIEPDVGV
     501  RGNVVLNLEK DLARSLGVAS IRVETIPGK TCMGLELPNP KRQMIRLSEI
     20  551  FNSPEFAESK SKLTLALGQD ITGQPVVTDL GKAPHLVAG TTGSGKSVGV
     601  NAMILSMLFK AAPEDVRMIM IDPKMLELSI YEGITHLLAP VVTDMKLAAN
     651  ALNWCVNEME KRYRLMSFMG VRNLAGFNQK IAEAAARGEK IGNPFSLTPD
     701  DPEPLEKLPF IVVVVDEFAD LMMTAGKKIE ELIARLAQKA RAAGIHLILA
     751  TQRPSVDVIT GLIKANIPTR IAFQVSSKID SRTILDQMG AENLLGQDML
     25  801  FLPPGTAYPQ RVHGAFASDE EVHRVVEYLK QFGEPTYDD ILSGGGSEEL
     851  PGIGRSGDGE TDPMYDEAVS VVLKTRKASI SGVQRALRIG YNRAARLIDQ
     901  MEAEGIVSAP EHNGNRTILV PLDNA*

```

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng (SEQ ID NO: 494) and FtsK (accession number p46889) (SEQ ID NO: 1142) show a 65 % amino acid identity in 459 overlap:

```

35      ORF58ng:  467  IEEKLAEFKVKVKVVDYSYSGPVITRYEIEPDVGVRGNSVLNLEKDLARSLGVASIRVVET 526
      FtsK:       868  +E +LA+F++K VV+  GPVITR+E+  GV+  + NL +DLARSL ++RVVE
      VEARLADFRKADVVNYSPPGPVITRFELNLAPGVKAARISNLSRDLARSLSTVAVRVVEV 927

      ORF58ng:  527  IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTLALGQDITGQPVVTDLGKAPHL 586
      IPGK  +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
      FtsK:       928  IPGKPYVGLELPNKKRQTVYLREVLDNAKFRDNPSPLTVVLGKDIAGEPVVADLAKMPHL 987

40      ORF58ng:  587  LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVVTDMDK 646
      LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK
      FtsK:       988  LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVVTDMDK 1047

      ORF58ng:  647  LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP-- 704
      AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA  I +P+  D  +
45      FtsK:       1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYWKPGDSMDAQH 1107

```

ORF58ng: 705 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 762  
 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL  
 FtsK: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167

ORF58ng: 763 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLPPGTAYPQVRVHGAFASDEEV 822  
 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV  
 FtsK: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227

ORF58ng: 823 HRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 882  
 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG  
 FtsK: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286

ORF58ng: 883 VQRALRIGYNRAARLIDQMEAEIGIVSAPEHNGNRTILVP 921  
 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P  
 FtsK: 1287 VQRQFRIGYNRAARIIEQMEAQIVSEQHNGNREVLAP 1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be (SEQ ID NO: 495):

1 ATGTTTGTGA TAGTTTGTAT CGTTATgtg TTGCTTGC GC TTGCCGGCCT  
 51 GTTTTTTGTG CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG  
 101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAA  
 151 GACGGTATGC CCGATTTTCC CGAGTTTCC CTGATGCTTT TCCATGCCGT  
 201 CAAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA  
 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT  
 301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGGTATT CAGACAGTGG  
 351 AAACGGGACG GAAGAAGCGG AAACGGAAGC AGCAGAAGCT GCGAGGAAG  
 401 AGGCTGCCgA TACgGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC  
 451 ATCCcatTCG ACCCGAGTAT TGCTGAAGGG TTGATGCAGT CTGAAAGCAA  
 501 AACTTCGCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA  
 551 CGCGTGCTTT AAGCAGCGCG GCTTAAAGGG AAACGAAAAA ACGCTATATC  
 601 GATGCATTTG AGAAAAACGG AACAGCCGTC CCCAAAGTAC GCGTGTCGGA  
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC  
 701 AACGCACGTA TTCCCGTATG TTTGATGCGG ACAAAGAAGC GTTTTCCGAG  
 751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC  
 801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC  
 851 ATGCAGGCA GGAGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC  
 901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCCGCCGCCG  
 951 CGTTTCCGTC AATTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG  
 1001 CGCGGATTTG CGCCCTGATT CCGGAAAGTC GGACGGTTGT CCGGAAACGG  
 1051 GATGTCGAAA TGCCGTCTGA AACCGAAAT GTTTTCACGG AAACCGTTTC  
 1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAGCTGCC GATATCCATA  
 1151 TTGAAGAGCC TGCCGCGCCC GATGCTTGGG TGGTCAAGC ACCCGAAGTG  
 1201 CCGGAGGTAG CCGTACCCGA AATCGATATT CTCCGCCGCG CTCCCGTATC  
 1251 GGAAATCTAC AACCGTACCT ATGAGCCGCC GGCAGGATTG GAGCAGGCGC  
 1301 AACGCAGCCG CATTGCCGAA ACCGACCATC TTGCCGCTGA TGTTTGAAT  
 1351 GGAGGTGGC AGGAGGAAAC CGCCGCTATT GCAGATGACG GCAGTGAGGG  
 1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAACC GAAGCGTTTCG  
 1451 GGCATGACAG TCAGGCGGTT TGTCCGTTG AAGATGTGCC GTCTGAACGC  
 1501 CCGTCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC  
 1551 GGAAGAGACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC  
 1601 TGCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG  
 1651 GAAAACAGCA TCACCATCGA AGAAAATTG GCGGAGTTCA AAGTCAAGGT  
 1701 CAAGGTGTC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG  
 1751 AACCCGATGT CGGCGTGCGC GGCAATTCCG TTCTGAATTT GGAAAAAGAC  
 1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCC  
 1851 CGGCAAAACC TGCAATGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA  
 1901 TACGCTGAG CGAAATTTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC  
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATT ACCGACAGC CCGTCGTAAC

5  
10  
15  
20

```

2001 CGACTTGGGC AAAGCACCGC ATTTGCTGGT TGCCGGCAGC ACCGGTTCGG
2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTCAAAGCC
2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151 GAGCATTAC GAAGGCATCA CGCACCTGCT CGCCCCTGTC GTTACCGATA
2201 TGAAGCTGGC GGCAAACGCG CTGAACTGGT GTGTTAACGA AATGGAAGAA
2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGCTTCAA
2301 CCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAAATC GGCAATCCGT
2351 TCAGCCTCAC GCCCGACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC
2401 GTGGTCGTGG TCGATGAGTT TGCCGATTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCGC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501 TCCACCTTAT CCTTGCCACA CAACGCCCCA GCGTCGATGT CATCACGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2601 AATCGACAGC CGCACGATTC TCGACCAAAT GGGCGCGGAA AACCTGCTCG
2651 GTCAGGCGCA TATGCTGTTT CTGCCCGCG GTACTGCCTA TCCGCAGCGC
2701 GTTCACGGCG CGTTTGCCTC GGATGAAGAG GTGCACCGCG TGGTCAATA
2751 TCTGAAGCAG TTTGGCGAGC CGGACTATGT TGACGATATT TTGAGCGGCG
2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGCGAAACC
2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC
2901 CAGCATTTCG GGCCTACAGC GCGCCTTGCG CATCGGCTAC AACCGCGCCG
2951 CGCGTCTGAT TGACCAAATG GAAGCGGAAG GCATTGTGTC CGCACCGGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 496; ORF58ng-1):

25  
30  
35  
40  
45

```

1 MFWIVLIVIV LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK
51 DGMPDFPEFS LMLFHAVKTA VYWLFVGVVR FCRNYLAHES EPDRPVPPAS
101 ANRADVPTAS DGYSDSGNGT EEAETEAAEA AEEEEADTED IATAVIDNRR
151 IPFDRSIAEG LMQSEKSTSP VRPVFKEITL EEATRALSSA ALRETKKRYI
201 DAFEKNGTAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE
251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQKEG QAEAKSPDVS
301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
351 DVEMPSETEN VFTETVSSVG YGGPVYDEAA DIHIEEPAAP DAWVVEPPEV
401 PEVAVPEIDI LPPPPVSEIY NRTYEPPAGF EQAQRSRIAE TDHLAADVLN
451 GGWQEETAAI ADDGSEGAEE RSSGQYLSET EAFGHDSQAV CPFEDVPSE
501 PSCRVSDETA DEGAFOQSEET GAVSEHLPTT DLLLPPLFNP EATQTEEELL
551 ENSITIEEKL AEFKVKVKV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD
601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
651 KLTLALQDI TGQPVVTDLG KAPHLVAGT TSGSKSVGVN AMILSMLFKA
701 APEDVRMIMI DPKMLELSIY EGITHLLAPV VTDMLAANA LNWCVNEMEK
751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTDD PEPEKLPFI
801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LPPGTAYPQR
901 VHGAFADEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDGET
951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
1001 HNGNRTILVP LDNA*

```

ORF58ng-1 (SEQ ID NO: 496) and ORF58-1 (SEQ ID NO: 490) show 97.2% identity in 1014 aa overlap:

50

```

          10      20      30      40      50      60
orf58-1.pep MFWIVLIVILLALAGLFFVRAQSEREWREVS AWQEKKG EKQAELEPEIK DGMPDFPELA
          |||:|||||
orf58ng-1 MFWIVLIVIVLLALAGLFFVRAQSEREWREVS AWQEKKG EKQAELEPEIK DGMPDFPEFS
          10      20      30      40      50      60

          70      80      90     100     110     120
orf58-1.pep LMLFHAVKTAVYWLFVGVVRFCRNYLAHES EPDRPVPPAS ANRADVPTAS DGYSDSGNGT

```



	orf58ng-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDRPVPPASANRADVPPTASDGYSDSGNGT 70 80 90 100 110 120
5	orf58-1.pep	EEAETEEAAEEEEAADTEDIAVAIDNRRIPFDRSIAEGLMPSESEISPVRPVFKEITL 130 140 150 160 170 180
	orf58ng-1	EEAETEAAEAEEEEAADTEDIAVAIDNRRIPFDRSIAEGLMQSESKTSPVRPVPFKEITL 130 140 150 160 170 180
10	orf58-1.pep	EETRALNSAALRETKKRYIDAFEKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSHM 190 200 210 220 230 240
	orf58ng-1	EETRALSSAALRETKKRYIDAFEKNGTAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSRM 190 200 210 220 230 240
15	orf58-1.pep	FDADKEAFSESADYGFEYPFEKQHPSAFSAVKAENARNAPFHRHAGQGKGQAEAKSPDVS 250 260 270 280 290 300
	orf58ng-1	FDADKEAFSESADYGFEYPFEKQHPSAFSAVKAENARNAPFRRHAGQEKGQAEAKSPDVS 250 260 270 280 290 300
20	orf58-1.pep	QGQSVDGTA VRDARRRVSVNLKEPNKATVSAEARISRLIPESQT VVGKR DVE MPSETEN 310 320 330 340 350 360
	orf58ng-1	QGQSVDGTA VRDARRRVSVNLKEPNKATVSAEARISRLIPESRT VVGKR DVE MPSETEN 310 320 330 340 350 360
25	orf58-1.pep	VFTETVSSVGYGGPVYDETADIHIEEPAAPDAWVPEPPEVKVPMTAIDIQPPPPVSEIY 370 380 390 400 410 420
	orf58ng-1	VFTETVSSVGYGGPVYDEAADIHIEEPAAPDAWVPEPPEVAVPEIDILPPPPVSEIY 370 380 390 400 410 420
30	orf58-1.pep	NRTYEPPSGFEQVQRSRIAETHLADDVLNGGWQEETA AIAD DGSEGAAERS SSGQYL SET 430 440 450 460 470 480
	orf58ng-1	NRTYEPPAGFEQAQRSRIAETHL AADV L NGGWQEETA AIAD DGSEGAAERS SSGQYL SET 430 440 450 460 470 480
35	orf58-1.pep	EAFGHDSQAVCPFENVPSERPSCRVS DTEADEGA FP SEETG AVSEHL PTTDLLLPPLFN P 490 500 510 520 530 540
	orf58ng-1	EAFGHDSQAVCPFEDVP SERP SCRVS DTEADEGA FQ SEETG AVSEHL PTTDLLLPPLFN P 490 500 510 520 530 540
40	orf58-1.pep	EATQTEEELLENSITIEEK LAEFKV KVVVDS YSGPV IT RYEIEPD VGVR GNSVL NLEKD 550 560 570 580 590 600
	orf58ng-1	EATQTEEELLENSITIEEK LAEFKV KVVVDS YSGPV IT RYEIEPD VGVR GNSVL NLEKD 550 560 570 580 590 600
45	orf58-1.pep	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLT LALGQDI 610 620 630 640 650 660
	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLT LALGQDI 610 620 630 640 650 660

-372-

		670	680	690	700	710	720
	orf58-1.pep	TGQPVVTDLGKAPHL	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY				
5	orf58ng-1	TGQPVVTDLGKAPHL	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY				
		670	680	690	700	710	720
	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
10	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
15	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLPGTAYPQR					
20	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDDETPMYDEAVSV					
25	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEPTYDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
30	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

Furthermore, ORF58ng-1 (SEQ ID NO: 496) shows significant homology to the *E.coli* protein FtsK (SEQ ID NO: 1142):

35	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK )gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] )gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] )gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329 Score = 576 bits (1469), Expect = e-163 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)
40	Query: 556 IEEKLAEFKVKVKKVVDYSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE Sbjct: 868 VEARLADFRIKADVNNYSPGPVITRFELNLAGVKAARISNLSRDLARSLSTVAVRVVEV 927
45	Query: 616 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 675 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL Sbjct: 928 IPGKPYVGLELPNKKRQTVYLVREVLNDAKFRDNPSPLTVVLGKDIAPEGVVADLAKMPHL 987
	Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVVTDMK 1047

Query: 736 LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPTDDPEP-- 793  
 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +  
 Sbjet: 1048 DAANALRWCVNEMERRYKLSALGVRNLAGYNEKIAEADRMMPPIPDYWKPGDSMDAQH 1107

5 Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 851  
 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL  
 Sbjet: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167

Query: 852 IKANIPTRIAFQVSSKIDSRITLDQMGAEENLLGQGDMLFLPPGTAYPQRVHGAFASDEEV 911  
 IKANIPTRIAF VSSKIDSRITLDQ GAE+LLG GDML+ P + P RVHGAF D+EV  
 Sbjet: 1168 IKANIPTRIAFTVSSKIDSRITLDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227

10 Query: 912 HRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971  
 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG  
 Sbjet: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286

Query: 972 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 1010  
 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P  
 15 Sbjet: 1287 VQRQFRIGYNRAARIIEQMEAQIVSEQGHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 59

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 497):

20 1 ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG  
 51 CATTTTCGTC GTCCTCTTGG CCGTATTGGT CTCCACGCAG GCAATCAACC  
 101 TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG  
 151 GCATTGGTCG GCTTCTGGGT C.....  
 //

25 901 .....A TTGCCATCGG TTTGTTTTTA ATTTACCAA ACAGGCTGAC  
 951 CCTGCTTTT GAAGCCGTGG AAGACGGCAA AATCCATTT TGGCTCGGAC  
 1001 TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC  
 1051 GTCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT  
 1101 GACATTGAAA GCGGAAAAAT GA

This corresponds to the amino acid sequence (SEQ ID NO: 498; ORF101):

35 1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL  
 51 ALVGFVW.....  
 //

301 ...IAIGLFL IYQNLTLF EAVEDGKIHF WLGLLPMHII MFVLALILLR  
 351 VRSMPSQPFV QAVGKSLTLK G GK\*

Further work revealed the complete nucleotide sequence (SEQ ID NO: 499):

40 1 ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG  
 51 CATTTTCGTC GTCCTCTTGG CCGTATTGGT CTCCACGCAG GCAATCAACC  
 101 TGCTCGGCCG TGCCGCCGAC GGGCGTGTCT CCATCGATGC CGTGTGGCA  
 151 TTGGTCGGCT TCTGGGTCAT CCGTATGACG CCGCTTTTGC TGGTGTGAC  
 201 CGCATTTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CGCGACACGG

-374-

```

251 AAATGTCGGT CTGGCTATCC TCGGATTGG CATTGAAACA ATGGATACGC
301 CCGGTGATGC AGTTTGCCGT GCCGTTTGCC GTTTTG GTTG CCGTCATGCA
351 GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
401 TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
5 451 AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTCGAAA CCTTCGATAC
501 CGAATCCGGC ATCATGAAAA ACCTGTTCCT GCGCGAACAG GACAAAAACG
551 GCGGCGACAA CATCATCTTC GCCAAAGAAG GTAAC TTCTC GCTGAACGAC
601 AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
10 651 CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAACTC AACCTGATTA
701 TCAGCACCAC GCCCAAATC ATCGACCCCG TTTCCACCG CCGTACCATT
751 CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCTGCTTG
851 CCGTGCCGCT TTCCTATTTT AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGTT TTTAATTTAC CAAACCGGC TGACCCTGCT
15 951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
1001 CTATGCACAT TATCATGTTT GCCGTGCAC TCATCCTGTT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGGCGGA AAATGA

```

20 This corresponds to the amino acid sequence (SEQ ID NO: 500; ORF101-1):

```

1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
51 LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPFA VLVAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
25 151 SLGKRNGRVY FVETFDTEG IMKNLFLREQ DKNGGDN IIF AKEGNFSLND
201 NKRTLELRHG YRYS GTPGRA DYNQVSFQKL NLIISTPKL IDPVSHRRTI
251 PTAQLIGSSN PQHQAE LMWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIMF AVALILLRVR
351 SMPSPFWQA VGKSLTLKGG K*

```

30 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 (SEQ ID NO: 498) shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) (SEQ ID NO: 502) from strain A of *N. meningitidis*:

```

35 orf101.pep      10      20      30      40      50
      |||||
orf101a          10      20      30      40      50
      |||||
//
40 orf101.pep      90      100     110
      .....IAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL
orf101a          280     290     300     310     320     330
      LTVSVLLLCLLAVPLSYFNPRSGHTYNILXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGL
45 orf101.pep      120     130     140     150
      LPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGKX
      |||||:|:|:|
orf101a          340     350     360     370
      LPMHIIMFVIAIVLLRVRSMPSQPFWQAVGKSLTLKGGKX
50

```

The complete length ORF101a nucleotide sequence (SEQ ID NO: 501) is:

	1	ATGATTATATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCGTCGCG
	51	CATTTTCGTC	GTCCCTCTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
5	101	TGCTCGGCCN	TGCCGCCGAC	NGGCGTNTCG	CCATCGATGC	CGTGTTGGCA
	151	TGGTCTGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTTCG	TNGTGTTGAC
	201	CGCATTTATC	AGTACGTTGA	CCGTGTTGAC	CCGCTACTGG	CNGNACAGCG
	251	AAATGTCGGT	CTGGNTATCC	TGCGGATTGG	CATTGAAACA	ATGGATACGC
	301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTTGGTTG	CCGTCATGCA
10	351	GC'TTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
	401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTTG	TGGAGGCAGG	CGGGTTC AAC
	451	AGTTTGGGCA	AGCGCAACGG	CAGGTTTAT	TTTGTGAA	CCTTCGATAC
	501	CGAATCCGGC	ATCATGAAAA	ACCTGTTCT	GCGCGAACAG	GACAAAAACG
	551	GCGGCGACAA	CATCATCTTC	NCCAAAGAAA	GTA ACTTCTC	GCTGAACGAC
15	601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCC
	651	CGGACGCGCC	GA CTACAATC	AGGTTCCTT	CCNAAAATC	AACTGATTA
	701	TCAGCACCAC	GCCCAAATC	ATCGACCCCG	TTTCCCACCG	CCGTACNATN
	751	CCNACNGCCC	AACTGATTGG	CAGCAGCAAC	CCGCAACTC	ANGCGGAATT
	801	GATGTGGCGC	ATCTCGCTGA	CCGT CAGCGT	CCTCTACTC	TGCGTGCTTG
	851	CCGTGCCGCT	TTCCTATTTC	AACCCGCGCA	GCGGACATAC	CTACAATATC
20	901	TTGANTGCCA	TCGGTTTGTT	TTTAATTTAC	CAAAACGGGC	TGACCCTGCT
	951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGGCTC	GGACTGCTGC
	1001	CTATGCACAT	CATCATGTTT	GT CATCGCAA	TCGTACTTCT	GCGCGTCCCG
	1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT
	1101	GAAAGCGCGA	AAATGA			

25 This encodes a protein having amino acid sequence (SEQ ID NO: 502):

30

```

      1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGXAAD XRXAIDAVLA
     51 LVGFVWXXMT PLLLVLTAFI STLTVLTRYW RDSEMSVWXS CGLALKQWIR
    101 PVMQFAVPFA VLVAVMQLWV IPWAELSRSRE YAEILKQKQE LSLVEAGGFN
    151 SLGKKNRGRVY FVETPDTEG IMKNLFLREQ DKNGGDNIIIF XKESNFSLND
    201 NKRTLELRHG YRYSGTPGRA DYNQVSFXKL NLIISTTPKL IDPVSHRRTX
    251 PTAQLIGSSN PQHXAEMLMW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
    301 LXAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIMF VIAIVLLRVR
    351 SMPSQPFWQA VGKSLTLKGG K*
```

35 ORF101a (SEQ ID NO: 502) and ORF101-1 (SEQ ID NO: 500) show 95.4% identity in 371 aa overlap:

[illegible]



```

5   351 GCTTTGGGTG ATACCGTGGG CAGAGCTGCG CAGCCGCGAA TATGCCGAAA
    401 TTTTGAAGCA GAAGCAGGAA TTGTCTTTGG TGAAGCCGG CGAGTTCAAT
    451 AACTTGGGCA AGCGCAACGG CAgggtttaT TtcgtcgaaA CCTTTGACAC
    501 CGaatccgGC ATCATGAAAA ACCTGTtcct GcGCGAACAG GACAAAAACG
    551 gcggcgacaA CATCATCTTC GCcaaaGAag gtaactTctc gctgaaggac
    601 AACAAAcgca cgctcgAATT GCGCCACGGC TACCGTTACA GCGGcacgcC
    651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAcTc aacctgATta
    701 TCAGCACCAC GCCCAAcTT ATCGaccCCG TTTCCACCG CCGCACCATT
    751 tcgacCGCCC AAcTGATGG CAGCAGCAAT CCGCAACATC AGGCAGAATT
10  801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCTGCTC TGCCTACTCG
    851 CCGTGCCGCT TTCCTATTTT AACCCGCGCA GCGGACATAC CTACAATATC
    901 TTGATTGCCA TCGGTTTGTT TTTAATTTAC CAAACGGGC TGACCCTGCT
    951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
15 1001 CTATGCACAT CATCATGTTC GTCATCGCAA TCGTACTTCT GCGCGTCCGC
    1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
    1101 GAAAGgcgGA AAATGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 506; ORF101ng-1):

```

20  1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
    51 LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
    101 PVMQFAVPFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
    151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNII F AKEGNFSLKD
    201 NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
    251 STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
25  301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIMF VIAIVLLRVR
    351 SMPSQPFWQA VGKSLTLKGG K*

```

ORF101ng-1 (SEQ ID NO: 506) and ORF101-1 (SEQ ID NO: 500) show 97.6% identity in 371 aa overlap:

```

30  orf101-1.pep  10      20      30      40      50      60
    MIYQRNLIKELSFTAVGIFV VLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT
    orf101ng-1   10      20      30      40      50      60
    MIYQRNLIKELSFTAVGIFV VLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWVIGMT

35  orf101-1.pep  70      80      90      100     110     120
    PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV
    orf101ng-1   70      80      90      100     110     120
    PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFA ILIAVMQLWV

40  orf101-1.pep  130     140     150     160     170     180
    IPWAE LRSREYAEILKQKQE LSLVEAGEFN SLGKRNGRVYFVETFDTESGIMKNLFLREQ
    orf101ng-1   130     140     150     160     170     180
    IPWAE LRSREYAEILKQKQE LSLVEAGEFN NLGKRNGRVYFVETFDTESGIMKNLFLREQ

45  orf101-1.pep  190     200     210     220     230     240
    DKNGGDNII F AKEGNFSLNDNKRTLELRHG YRYSGTPGRADYNQVSFQKL NLIISTTPKL
    orf101ng-1   190     200     210     220     230     240
    DKNGGDNII F AKEGNFSLKDNKRTLELRHG YRYSGTPGRADYNQVSFQKL NLIISTTPKL

50  orf101-1.pep  250     260     270     280     290     300
    IDPVSHRRTIPTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI

```

	orf101ng-1	 IDPVSHRRITSTAQLIGSSNPQHQAELMWRLSLTVSVLLCLLAVPLSYFNPRSGHTYNI	250	260	270	280	290	300
5	orf101-1.pep	310 320 330 340 350 360 LIAIGLFLIYQNGLTLLFEAVEDGKIHFGLLPMHIIMFAVALILLRVRSMPSQPFWQA						
	orf101ng-1	 LIAIGLFLIYQNGLTLLFEAVEDGKIHFGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	310	320	330	340	350	360
10	orf101-1.pep	370 VGKSLTLKGGKX						
	orf101ng-1	 VGKSLTLKGGKX						
		370						

- 15 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 60

- 20 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 507):

1 ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC  
51 GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG  
101 TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT  
25 151 ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT  
201 TCGTGTCTGC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT  
251 CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC  
301 AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG  
351 TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCTGA GCAAGCAGGC  
30 401 ATTCGTAA

This corresponds to the amino acid sequence (SEQ ID NO: 508; ORF113):

1 ..GGGFINASCA TLTTAKPOYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR  
51 ILSYHSKIDA PVWGQDVRV AGQNDVAATG DAHSPILNNA AANTSNNATAN  
35 101 NGTHIPLFAI DTGKLGGXVC QQNHLDQYGR ASRHS\*

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein (SEQ ID NO: 1143) of *N.meningitidis* (accession AF030941)

ORF (SEQ ID NO: 508) and pspA (SEQ ID NO: 1143) show 44% aa identity in 179aa overlap:



```

orf113  GGGFINASCATLTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
        GGG INA+ TLT+ P   G+L+ F + G VVI G GLD D DYTRILS ++I+A
pspa    GGGLINAASVTLTSGVPVLNNGNLTGFDVSSGKVIGGKGLDTSADYTRILSRAAEINA 256

orf113  PVWQDVRVVGQNDVAATGDAHSPILXXXXXXXXXXXXXXXXGTHIPLFAIDTGKLGMYA 120
        VWG+DV+VV+G+N +   G                               + P AIDT LGGMYA
pspa    GVWGKDVKVVS GKNKLD F DG-----SLAKTASAPSSSDSVTPTVAIDTATLGMYA 307

orf113  NKITLISTVEQAGIRNQGWFFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
        +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A       +++ A+ V N
pspa    DKITLISTDNGAVIRNKGRIFAATGGVTL S ADGKLSNSGSIDAA----EITISAQTVDN 362

```

### 10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 (SEQ ID NO: 508) shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa overlap at the C-terminal part with a predicted ORF (ORF113ng) (SEQ ID NO: 510) from *N. gonorrhoeae*:

```

15  orf113                               GGGFINASCATLTAKPQYQAGDLSAFKIR 30
        ||||| |||:|:|:|:|:|:|:|:|
orf113ng SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQAGDFSGFKIR 224

orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWQDVRVVGQNDVAATGDAHSPILNNA 90
        ||:|||||:||||:|
orf113ng QGNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263

20  orf113                               IDTGKLGXVCQQNHLDQYGRASRHS 135
        ||||| |||:|:|:|:|:|
orf113ng DFSGFKIRQGNNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263

```

The complete length ORF113ng nucleotide sequence (SEQ ID NO: 509) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 510):

```

25      1 MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
      51 SKAFCF SALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGN GIP
      101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
      151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
      201 ASRATLTGQ PQYQAGDFSG FKIRQGNNAVI AGHGLDARDT DFTRILVCQQ
30      251 NHLDQYGRTS RHS*

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 61

35 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 511):

```

1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG

```

51 CAACATTTCA CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTA  
 101 GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT  
 151 TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT  
 201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC  
 5 251 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGACAGC  
 301 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 351 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 401 GTTTAGAcGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT  
 451 AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGGCATT  
 10 501 AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 551 AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA  
 601 CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT  
 651 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT  
 701 CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 15 751 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 801 ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT  
 851 TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT  
 901 CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 951 TATCACAGGC AAAGAAAAAG GTGTTT..

This corresponds to the amino acid sequence (SEQ ID NO: 512; ORF115):

1 ..STGHSEQNYT LPREITRNI LGSFAYESHR KALSHHAPSQ GTELPQSNIGI  
 51 SLPYTSNSFT PLPSSSLYII NPVNKGYLE TDPRFANYRQ WLGS DYMLDS  
 25 101 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 151 NGATAARSMN LSVGIALSAE QVAQLTSDIV WLQKEVKLP DGGTQTVLVP  
 201 QVYVRVKNKD IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 251 DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTLNAGNN INSQSTASS  
 301 QNTQGSSTYL DRMAGIYITG KEGV..

30 Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein (SEQ ID NO: 1143) of *N.meningitidis* (accession number AF030941)

ORF115 (SEQ ID NO: 512) and pspA protein (SEQ ID NO: 1143) show 50% aa identity in 325aa overlap:

35 Orf115: 1 STGHSEQNYTLPREITRNI LGSFAYESHRKALSHHAPSQGTELPQSNIGISLPYTSNSFT 60  
 STG+S Y E++ +I +G AY+ + + P + NGI +T  
 pspA: 778 STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVVPVVAENGIHPTFT----- 831

Orf115: 61 PLPSSSLYIINPVNKGYLETDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQR 120  
 LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQR+  
 40 pspA: 832 -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSGYMLAALQDDPNHIHKRLGDGYEQR 890

Orf115: 121 LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180  
 L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV  
 pspA: 891 LVNEQIAKLTGYRRLDGYTNDEEQFKALMDNGITI AKELQLTPGIALSAEQVARLTSDIV 950

Orf115: 181 WLQKEVKLPDGGTQTVLVPQVYVRVKNKDIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239  
 WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG  
 45 pspA: 951 WLENETVTLPDGTQTVLKPQVYVRARPKDMNGQALLSGSVVDIG-SGA IENRGGLIAG 1009

Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXXXX 299  
 R ALI+N + N+ G + + A DI N G + AE LLL A  
 pspA: 1010 REALILNAQNIKNLQGDLOGKNIFAAAGSDITNTGS-IGAENALLLKASNIESRSETRS 1068

Orf115: 300 XXXXXXXXXYLDRMAGIYITGKEKG 324  
 + R+AGIY+TG++ G  
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF115 (SEQ ID NO: 512) shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) (SEQ ID NO: 514) from *N.gonorrhoeae*:

10	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
	orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESHK	71
	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVE	81
15	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET	131
	orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	191
20	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLQKEVKLPDGGTQTVLVPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNKGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
	orf115ng	VYVRVKNKGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
25	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
	orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
	orf115.pep	EKGV	325
30	orf115ng	EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFDADNHTIR	431

An ORF115ng nucleotide sequence (SEQ ID NO: 513) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 514):

35	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPTEEITRDIS	LGSFAYESHS	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGSDYMLGS
	151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLQKEVKLP	DGGTQTVLMP
	251	QYVVRVKNKG	IDGKGALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
40	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIIAGQ	ISNQSDQGQT
	401	RLQAGRDINL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGSS	IQTKGDVTL
	451	SGNNLNAAK	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG

501 GNKLVIDDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAC QFDKAKTTAL  
 701 MPWRLPMQVG RLFKQAKAPK K\*

Further work revealed the following partial gonococcal DNA sequence (SEQ ID NO: 515):

10 1 TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG  
 51 CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACAAAC TACTGGCGTG  
 101 CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT  
 151 TTGCCGGAGG AAATCACACG CGACATTTCA CTGGGTTTCAT TTGCCTATGA  
 201 ATCGCATAGC AAAGCATTAA GCCGTCAATG GCCCAGCCAA GGCACGTAGT  
 251 TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAAG CAACGGTATT  
 301 TCGCTACCCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT  
 15 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC  
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC  
 451 CTCAAAC TAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT  
 20 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGTCAAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT  
 25 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT  
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT  
 30 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGCAA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 35 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 40 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 45 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTG AGTTCGCCCG TTACCGATTG GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA  
 50 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence (SEQ ID NO: 516; ORF115ng-1):

55 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS  
 151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD

5  
10  
201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP  
251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGNN INNSTAKSS  
351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIIAGQ ISNQSDQGQT  
401 RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
451 SGNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
501 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI  
551 QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS  
601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ  
651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
701 MPWRLPMQVG RPIKQAKAHK T\*

This gonococcal protein (ORF115ng-1) (SEQ ID NO: 516) shows 91.9% identity with ORF115 (SEQ ID NO: 512) over 334aa:

15  
20 30 40 50 60 70  
orf115ng-1.p NEQTFGEKKVSENGKLHNYWRARRKGHDETHREQNYTLPEEITRDISLGSFAYESHK  
orf115 STGHSEQNYTLPREITRNISLGSFAYESHK  
10 20 30

20  
80 90 100 110 120 130  
orf115ng-1.p ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET  
orf115 ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLET  
40 50 60 70 80

25  
140 150 160 170 180 190  
orf115ng-1.p DPRFANYRQWLGS DYMLGSLKLDPNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND  
orf115 DPRFANYRQWLGS DYMLDSLKLDPNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND  
90 100 110 120 130 140

30  
200 210 220 230 240 250  
orf115ng-1.p EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVLVQKEVKLPDGGTQTVLMPQ  
orf115 EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVLVQKEVKLPDGGTQTVLVPQ  
150 160 170 180 190 200

35  
260 270 280 290 300 310  
orf115ng-1.p VYVRVKNNGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK  
orf115 VYVRVKNNGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK  
210 220 230 240 250 260

40  
320 330 340 350 360 370  
orf115ng-1.p SAVTATQDINNIGGILSAEQTLNAGNNINNSTAKSSQNAQGSSTYLDRMAGIYITGK  
orf115 SAVTATQDINNIGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK  
270 280 290 300 310 320

45  
380 390 400 410 420 430  
orf115ng-1.p EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADADNHTIR  
orf115 EKGV

In addition, it shows homology with a secreted *N.meningitidis* protein (SEQ ID NO: 1143) in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273

5      Score = 604 bits (1541), Expect = e-172  
      Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1      LLVQTEKDGLHNEQTFGEKKVFSENGKLNHYWRARRKGHDETGHRQNYTLPEEITRDIS 60  
              L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I  
Sbjct: 739    LIVGTPESALDNDDELGTGKTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796

10      Query: 61    LGSFAYESHKALSRRHAPSQGTETLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
              +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
Sbjct: 797    MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

15      Query: 121    NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEYQRLINEQIAELT 180  
              P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEYQ+L+NEQIA+LT  
Sbjct: 841    APNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEYQKLVNEQIAKLT 900

Query: 181    GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQEVKLP 240  
              G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP  
Sbjct: 901    GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTL P 960

20      Query: 241    DGGTQTVLMPQVYVRVKNNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299  
              DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N  
Sbjct: 961    DGTQTVLKPQVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300    LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359  
              + N+ G + + A D I N G I AE LLL A NNI ++S +S+QN QGS  
Sbjct: 1020    IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

25      Query: 360    LDRMAGIYITGKEKGVLAAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419  
              + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q  
Sbjct: 1079    IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

30      Query: 420    EIHFADNHTIRGSTNEVGSSIQTKGDVTLNAGNNLNAAAEVGSAGKTLAVYAKNDITI 479  
              FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +  
Sbjct: 1139    NTIFDSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV 1198

Query: 480    SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
              +G + +DA K+TGRSGG K +T ++ + A S T +GK+++L +G D + G  
Sbjct: 1199    EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

35      Query: 540    SNVISDNGTRIQAGNHVRIGTTQTQSQSEYHQTKSGLM-SAGIGFTIGSKTNTQENQS 598  
              SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
Sbjct: 1259    SNIIADNHTILSAKNNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318

Query: 599    QSNEHTGSTVGS LKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658  
              ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
Sbjct: 1319    ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISGKISIDAAQNRYSQESK 1378

40      Query: 659    QTYEQKGLTVAFSSPVT D 676  
              Q YEQKG+TVA S PV +  
Sbjct: 1379    QVYEQKGVTVVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 62

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 517):

```

5      1  ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
      51  TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
     101  ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
     151  GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
     201  CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTCAG GCAGGAAACG
10     251  ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
     301  CAAGCAGGCA ATCATGTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG
     351  CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
     401  GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
     451  AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
15     501  TGTTCGAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
     551  AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
     601  AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
     651  GGTGGCATTC AGTTCGCCCC TTACCGATTT GGCACAACAA ...

```

20 This corresponds to the amino acid sequence (SEQ ID NO: 518; ORF117):

```

      1  ..SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
     51  GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTQI
     101  QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS
     151  NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNTIY AQSIDIQAAH
25     201  NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein (SEQ ID NO: 1143) of *N.meningitidis* (accession number AF030941)

30 ORF117 (SEQ ID NO: 518) and pspA protein (SEQ ID NO: 1143) show 45% aa identity in 224aa overlap:

```

Orf117: 4  NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIDDKAQSH 63
      ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
pspA: 1173 DIRIRAAEVGSEQGRLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQ 1232
35 Orf117: 64 HETAQSSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123
      + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIADNHTILSAKNNIVLKAETRSRSAEMNKK 1292
Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNHTGSTVGS LKGDTTIVAGKHYEQIGSTVSS 182
      +KSGLM S GIGFT GSK +TQ N+S++ HT S VGSL G+T I AGKHY Q GST+SS
40 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

```

Orf117: 183 PEGNNTIYAQSIDIQAAHNKLSNNTTQTYEQKXLTVAFSSPVT 226  
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +  
 pspA: 1353 PQGDVGISGKISIDAAQNRYSESQVYEQKGVTVVAISVPVVN 1396

### Homology with a predicted ORF from *N.gonorrhoeae*

- 5 ORF117 (SEQ ID NO: 518) shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) (SEQ ID NO: 520) from *N.gonorrhoeae*:

	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFDADNHTIRGSTNEVGSSIQTKGDVTLTSGNNLNAAAEVGSAGKTLAVYAKNDITIS	480
10	orf117.pep	AGINTTHVDDASKHTGRSGGNGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
	orf117ng	SGIHAGQVDDASKHTGRSGGNGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
15	orf117ng	NVISDNGTRIAGNHVRIGTTQTQSQSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	600
	orf117.pep	NEHTGSTVGS LKGDTTIVAGKHYESQIGSTVSSPEGNNTIYAQSIDIQAAHNKLSNNTTQT	210
	orf117ng	NEHTGSTVGS LKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMIDIGAAQNQLNSKTTQT	660
20	orf117.pep	YEQKXLTVAFSSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVTDLAQQAI AVAHKAAKQFDKAKTTALMPWRLPMQVGRLEFKQAKAPK	720

An ORF117ng nucleotide sequence (SEQ ID NO: 519) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 520):

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
	151	LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
30	251	QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLNAGNN INNQSTAKSS
	351	QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINIAGQ ISNQSDQGQT
	401	RLQAGRDINL DTVQGTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLT
	451	SGNNLNAAAEVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
35	501	GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSETY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
	601	NEHTGSTVGS LKGDTTIVAS KHYESQSGNSV SSPEGNNLIS TQSMIDIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKA AKQFDKAKTTAL
40	701	MPWRLPMQVG RLFKQAKAPK K*

Further work revealed the following gonococcal partial DNA sequence (SEQ ID NO: 521):

	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACAAAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT



5  
10  
15  
20  
25  
30  
35  
40

```

151 TTGCCGGAGG AAATCACACG CGACATTTCA CTGGGTTTCAT TTGCCTATGA
201 ATCGCATAGC AAAGCATTAA GCCGTATGCG GCCCAGCCAA GGCCTGAGT
251 TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
301 TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGTA ACCGATCCAC
401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT
601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
801 GTTGTCAAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAGACA
1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGGCAA
1251 ATATCAAGAA ATCCATTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCctatTG
1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
1501 GGTAATAAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG
1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCAACA
1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
2001 GGTGGCATT CAGTTGCCCCG TTACCGATTT GGCACAACAA GCGATTGCCG
2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
2101 ATGCCATGGC GGCTGCCAAT GCAGGTGGC AGGCCTATCA AACAGGCAAA
2151 GCGCACAAAA ACTTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 522; ORF117ng-1):

45  
50  
55

```

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
51 LPPEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
151 LKLDPNNLHK RLGDGYEYEQ LINEQIAELT GHRRLDGYQN DEEQFKALMD
201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
251 QVYVRVKNGG IDKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLLNAGNN INNQSTAKSS
351 QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINI IAGQ ISNQSDQGQT
401 RLQAGRDINL DTVQTKGYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLTLL
451 SGNNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
501 GNKLVIDTKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNNGTRI
551 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNLIS TQSMIDGAAQ
651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
701 MPWRLPMQVG RPIKQAKAHK T*

```

ORF117ng-1 (SEQ ID NO: 522) shows the same 90% identity over a 230aa overlap with ORF117 (SEQ ID NO: 518). In addition, it shows homology with a secreted *N.meningitidis* protein (SEQ ID NO: 1143) in the database:

```

5      gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length =
      2273
      Score = 604 bits (1541), Expect = e-172
      Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

10      Query: 1   LLVQTEKDGLHNEQTFGEKKVFSENGKLNHYWRARRKGHDETGHREQNYTLPEEITRDIS 60
      L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
      Sbjct: 739   LIVGTPESALDNDNETLGTCTI-TDKGDLHRYHRHHKKGRDSTGYRSRSPYEPAPEVS-SIR 796

      Query: 61   LGSFAYESHKALSRRHAPSQGTELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
      +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
      Sbjct: 797   MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

15      Query: 121  NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELT 180
      P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT
      Sbjct: 841   APNNKGYLIETDPAFTDYRKWLGSGYMLAALQDPNHIHKRLGDGYEQKLVNEQIAKLT 900

      Query: 181  GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVVLVQKEVKLP 240
      G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
      Sbjct: 901   GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP 960

20      Query: 241  DGGTQTVLMPQVYVRVKNNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
      DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
      Sbjct: 961   DGTQTQVLKPKVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

      Query: 300  LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSSTAKSSQNAQGSSTY 359
      + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
25      Sbjct: 1020 IKNLQGDLDQGNIFAAAGSDITNTGSI-GAENALLKASNIESRSETRSNQNEQGSVRN 1078

      Query: 360  LDRMAGIYITGKEKGVLAQAQKDINIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419
      + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
      Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

30      Query: 420  EIHFDADNHTIRGSTNEVGSSIQTKGDVTLTSGNNLNAAAEVGSAGKTLAVYAKNDITI 479
      FD+DN+ IR NEVG+S+I+T+G+++L + ++ +AAEVGS +G L + A DI +
      Sbjct: 1139 NTIFSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV 1198

      Query: 480  SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
      +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
      Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDIITVTG 1258

35      Query: 540  SNVISDNGTRIAGNHVRIGTTQTQSQSETYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
      SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
      Sbjct: 1259 SNIIADNHTILSAKNIVLKAATRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRN 1318

      Query: 599  QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTT 658
      ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
40      Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

      Query: 659  QTYEQKGLTVAFSSPVT 676
      Q YEQKG+TVA S PV +
      Sbjct: 1379 QVYEQKGVTVAISVPVN 1396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 63

5 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 523):

```

1  ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA
51  CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC
151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC
10  201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGyCATGCGC AACCTGCAAG
251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGGCA TTATCGGCAA
351 CTCCGCCAC ACCGTTTCCG AACCCEAAAC CGGACATTCC GCAACGAAAC
401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAA ACCTGCAAAA
15  451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
501 CGTGCGCATC GACTTCATCT CCTAT...
```

This corresponds to the amino acid sequence (SEQ ID NO: 524; ORF119):

```

1  MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR
20  51  DGKPSGGSVM MPKPQPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
151 PLITLKELSK VELSWFDVRI DFISY...
```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 525):

```

25  1  ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA
51  CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG
101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC
151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCGGC
30  201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGCCATGCGC AACCTGCAAG
251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGGCA TTATCGGCAA
351 CTCCGCCAC ACCGTTTCCG AACCCEAAAC CGGACATTCC GCACCGAAAC
401 CTGCCGACGC GCCGGCAAAA CCTGCACCCG TTCCGCAAA ACCTGCAAAA
35  451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA
501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTGCGC
601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
651 CTATCAGGCA TTTATCGTGG GTATTGAGGC AGTCAGCCGC AACGGCATTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTGCGA
40  751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
45  1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
```

1251 ACCGGGCGGC AAAACCGCAT TGC GCCTGTT CTCCTAA

This corresponds to the amino acid sequence (SEQ ID NO: 526; ORF119-1):

```

5      1  MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
      51  DGKPSGG SVM MPKPQPAVKK TAKPQDPAMR NLQE QDAVYI AKQKQAKASP
     101  FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
     151  PLITLKELSK VELPWFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
     201  CTMDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
     251  QSMGGQTLHT DLA AFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
    10  301  AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
     351  MLLDIPHSPA GEKT FDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
     401  RTYVLARQSE MLKVGIEPPG KTALRLFS*

```

Computer analysis of this amino acid sequence gave the following results:

#### 15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 (SEQ ID NO: 524) shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) (SEQ ID NO: 528) from strain A of *N. meningitidis*:

```

20  orf119.pep  10      20      30      40      50      60
      MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGKPSGG SVM
      |||||:|||||
 orf119a      10      20      30      40      50      60
      MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM

25  orf119.pep  70      80      90      100     110     120
      MPKPQPAVKKTAKPQDPXMRNLQE QDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
      |||||:|||||
 orf119a      70      80      90      100     110     120
      MPKPQPAVKKTAKSQDPAMRNLQE QDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH

30  orf119.pep  130     140     150     160     170
      TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFVRFDFISY
      |||||:|||||
 orf119a      130     140     150     160     170     180
      TVPEPQTGHSAPKPADAPAKPVPQTPAKPLITLKELSKVELPWFVRFDFISYIALTE

35  orf119a      190     200     210     220     230     240
      AKELHALPRLSNRCRYQIVGCTMDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS

```

The complete length ORF119a nucleotide sequence (SEQ ID NO: 527) is:

```

40  1  ATGATTACA TCGTACTGTT CCTCGCGCC GTCCTGCGC TTGTCGCTA
      51  CAATATGTAT CAGGAAAACC AATACGCAA AAAAGTGCGC GACCAGTTCTG
     101  GGCAC TCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCCGC
     151  GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
     201  GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG
     251  AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
     301  TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA
     351  CTCCGCCAC ACCGTTCCCG AACC CCAAAC CGGACATTCC GCACCAAAAC
    45  401  CTGCCGACGC ACCGGCAAAA CCTGTTCCCG TTCCGCAAAC GCCGGCAAAA
     451  CCGCTGATTA CGCTCAAAGA GCTGTGGAAG GTCGAGCTGC CCTGGTTTGA

```

-391-

5  
10  
15

```

501 CGTGCGCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC
551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTGGGC
601 TGCACCATGG ACGACCATT CCAGATTGCC GAACCCATCC CGGGCATCCG
651 CTATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTGCGA
751 CACAGCATGG GCGGTACAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATTG GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTTTGAA GACGACGCGC CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGCA GCGGAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TCGCGCTGTT CTCCTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 528):

20  
25

```

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSVHR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVVLARQSE MLKVGIEPGG KTALRLFS*

```

30 ORF119a (SEQ ID NO: 528) and ORF119-1 (SEQ ID NO: 526) show 98.6% identity in 428 aa overlap:

35  
40  
45  
50

```

              10      20      30      40      50      60
orf119a.pep  MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSVHRDGKPSGGPVM
              |||||:|||||
orf119-1     MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSVHRDGKPSGGPVM
              10      20      30      40      50      60

              70      80      90      100     110     120
orf119a.pep  MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
              |||||:|||||
orf119-1     MPKPQPAVKKTAKPQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
              70      80      90      100     110     120

              130     140     150     160     170     180
orf119a.pep  TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE
              |||||:|||||
orf119-1     TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE
              130     140     150     160     170     180

              190     200     210     220     230     240
orf119a.pep  AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
              |||||:|||||
orf119-1     AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
              190     200     210     220     230     240

```

-392-

		250	260	270	280	290	300
	orf119a.pep	AFNRQVDFAFAHSMGGQTLHTDLAAFIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS					
	orf119-1	AFNRQVDFAFAHSMGGQTLHTDLAAFIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS					
5		250	260	270	280	290	300
	orf119a.pep	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
10		310	320	330	340	350	360
	orf119a.pep	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
		310	320	330	340	350	360
	orf119a.pep	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
	orf119-1	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
15		370	380	390	400	410	420
	orf119a.pep	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
	orf119-1	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
		370	380	390	400	410	420
	orf119a.pep	KTALRLFSX					
	orf119-1	KTALRLFSX					
		429					

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 (SEQ ID NO: 524) shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) (SEQ ID NO: 530) from *N.gonorrhoeae*:

	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGPSPGGSV	60
	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGPSPGGPVM	60
25	orf119.pep	MPKPQPAVKKTAKPDQPMRNLOEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRNLOEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH	120
	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY	175
30	orf119ng	TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE	180

The complete length ORF119ng nucleotide sequence (SEQ ID NO: 529) is:

	1	ATGATTTACA	TCGTACTGTT	CCTCGCCGCC	GTCCTCGCCG	TTGTCGCCTA
35	51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGCGC	GACCAGTTTCG
	101	GACACTCCGA	CAAAGATGCC	CTGCTCAACA	GCAAACCCAG	CCATGTCCGC
	151	GACGGCAAAC	CGTCCGGCGG	GCCAGTCATG	ATGCCGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	CCGGCCAAAC	CCCAAGACTC	CGCCATGCGC	AACCTGCAAG
	251	AACAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTCCCCG
40	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGAA	GAAATCGGCA	TTATCGGCAA
	351	CTCCGCCAC	ACCGTTTCCG	AACCCCAAAC	CGGACATTCC	GCACCGAAAC
	401	CTGCCGACGC	GCCGGCAAAA	CCCGTTCCCG	TTCCGCAAAC	GCCGGCAAAA
	451	CCGCTGATTA	CGCTCAAAGA	GCTGTGGAAG	GTCGAGCTGC	CCTGGTTTGA
	501	CGTGCGCTtc	gACTTCATCT	CCTATATCGC	GCTGACCGAA	GCCAAAGAAC
45	551	TGCACGCACT	GCCGCGCCTT	tccAACCGCT	GCCGCTACCA	GATTGTGGGC
	601	TGCACCATGG	ACGACCATT	CCAGATTGCC	GAACCCATCC	CGGGCATCCG

651 CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG  
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGCGGA CGCATTTCGA  
751 CAAAGCATGG GCGGTGAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA  
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG  
5 851 CCATCCATTT GGTTCGCGG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC  
901 GCCGTACGG GCGTGGGTTT CGTTTGGGAA GACGACGGCG CGTTCCACTA  
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG  
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT  
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA  
10 1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGTCAGTTG AACCTGAATC  
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTA  
1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA  
1251 ACCGGGCGGC AAAACCGCCC TGCCTGTT TTCATAA

15 This encodes a protein having amino acid sequence (SEQ ID NO: 530):

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR  
51 DGKPSGGPVM MPKPQPAVKK PAKPQDSAMR NLQEQDAVYI AKQKQAKASP  
101 FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK  
20 151 PLITLKELSK VELPWFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG  
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAF  
251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS  
301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS  
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSQQL NLNLVNDKME EVSTQWLKDV  
25 401 RTYVRLARQSE MLKVGIEPGG KTALRLFS\*

ORF119ng (SEQ ID NO: 530) and ORF119-1 (SEQ ID NO: 526) show 98.4% identity over 428 aa overlap:

		10	20	30	40	50	60
30	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGPVM					
	orf119-1	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV RDGKPSGGSVM					
		10	20	30	40	50	60
35	orf119ng	MPKPQPAVKKPAKPQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH					
	orf119-1	MPKPQPAVKKTAKPQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
		70	80	90	100	110	120
40	orf119ng	TVSEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDFISYIALTE					
	orf119-1	TVSEPQTGHSAPKPADAPAKPAPVQTPAKPLITLKELSKVELPWFDFISYIALTE					
		130	140	150	160	170	180
45	orf119ng	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240
50	orf119ng	AFNRQADAFQSMGGQTLHTDLAAFIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS					
	orf119-1	AFNRQVDAFAQSMGGQTLHTDLAAFIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS					
		250	260	270	280	290	300

		250	260	270	280	290	300
		310	320	330	340	350	360
	orf119ng	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
5	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA					
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf119ng	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
10	orf119-1	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVRLARQSEMLKVGIEPGG					
		370	380	390	400	410	420
		429					
	orf119ng	KTALRLFSX					
15	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 20 Example 64

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 531)

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTCCTCCA
25	101	TCGCCCTGAT	TTCAATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTyGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTTCCGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTT	GACCGGAATC	GGCATCGCGT
30	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence (SEQ ID NO: 532; ORF134):

	1	..ARHGTEFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
35	51	LVSVTERTKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
	101	SLVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAACLNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence (SEQ ID NO: 533):

40	1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTCGC	TTCTGACGAT
	51	GCTCGGCATC	ATCATCGGTA	TCGCGTCGGT	GGTTTCCGTC	GTCGCATTGG
	101	GCAATGGTTC	GCAGAAAAAA	ATCCTTGAAG	ACATCAGTTC	GATAGGGACG
	151	AACACCATCA	GCATCTTCCC	GGGGCGCGGC	TTCGGCGACA	GGCGCAGCGG
	201	CAGGATTAAA	ACCCTGACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA



-395-

5  
10  
15  
20

```

251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
351 TTTCGACGTG CGCGGACTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAACCA TTTGTTCAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
651 AGACAAATGCC AATACCCAGG TTGCGGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTCTTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCTC
801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
851 TGCTGGTGTC CGTTACCGAG CGCACCAAG AAATCGGCAT ACGGATGGCA
901 ATCGGCGCGC GGC GCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC GTCATCGGCG GTTTGGTTCG CGTGGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTC
1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCTGCCA ATAAAGCAGC CAACTCAAT CCGATAGACG
1151 CATTGCACA GGATTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 534; ORF134-1):

25  
30

```

1 MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSIGT
51 NTISIFPGRG FGDRRSRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFA DSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 (SEQ ID NO: 1144) of *E.coli* (accession number AE000189)

35 ORF134 (SEQ ID NO: 532) and o648 protein (SEQ ID NO: 1144) show 45% aa identity in 153aa overlap:

40

```

Orf134: 2 RHGTEDFFMNNSDXIRQIVESTTGTMKXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
          RHG +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EI
o648: 496 RHGKKDFFTWNMDGVLKTVEKTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREI 555

Orf134: 62 GIRMAIGARRGNIXQQFLIEAXXXXXXXXXXXXXXFNHFVTDFFPMDISAMSVI 121
          GIRMA+GAR ++ QQFLIEA F+ + + S ++++
o648: 556 GIRMAVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALL 615

Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
          A CST GI FG++PA AA+L+P+DALA++
o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648

```

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 (SEQ ID NO: 532) shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) (SEQ ID NO: 536) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF134a nucleotide sequence (SEQ ID NO: 535) is:

25	1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTTCG	TTCTGACGAT
	51	GCTCGGCATC	ATCATCGGTA	TCGCTTCGGT	TGTCCTCCGT	GTCGCATTGG
	101	GCAACGGTTC	GCAGAAAAAA	ATCCTTGAAG	ACATCAGTTC	GATAGGGGACG
	151	AACACCATCA	GCATCTTCCC	AGGGCGCGGC	TTCCGGCGACA	GGCGCAGCGG
	201	CAGGATTAAA	ACCTTGACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA
30	251	GCTACGTTGC	TTCCGCCACG	CCCATGACTT	CGACGGCGCG	CACCGTGACT
	301	TACCGCAATA	CCGACCTGAC	CGCTTCTTTG	TACGGTGTGG	GCGAACAATA
	351	TTTCGACGTG	CGCGGGCTGA	AGCTGGA AAC	GGGGCGGCTG	TTTGACGAAA
	401	ACGATGTGAA	AGAAGACGCG	CAGGTCGTCT	TCATCGACCA	AAATGTCAAAA
	451	GACAAACTCT	TTGCGGACTC	GGATCCGTTG	GGTAAAACCA	TTTTGTTCAG
35	501	GAACGCCCCC	TTGACCGTCA	TCGGCGTGAT	GA AAAAGAC	GAAAACGCTT
	551	TCGGCAATTC	CGACGTGCTG	ATGCTTTGGT	CGCCCTATAC	GACCGTGATG
	601	CACCAAATCA	CAGGCGAGAG	CCACACCAAC	TCCATCACCG	TCAAAATCAA
	651	AGACAATGCC	AATACCCAGG	TTGCCGAAAA	AGGGGTGACC	GATCTGCTCA
	701	AAGCGCGGCA	CGGCACGGAA	GATTTCTTCA	TGAACAAACG	CGACAGCATC
40	751	AGGCAGATAG	TCGAAAGCAC	CACCGGTACG	ATGAAGCTGC	TGATTTCTCT
	801	CATCGCCCTG	ATTTCAATTG	TAGTCGGCGG	CATCGCGGTG	ATGAACATCA
	851	TGCTGGTGTG	CGTTACCGAG	CGCACCAAAG	AAATCGGCAT	ACGGATGGCA
	901	ATCGGCGCGC	GGCGCGGCAA	TATTTTGCAG	CAGTTTTTGA	TTGAGGCGGT
	951	GTTAATCTGC	GTCATCGGCG	GTTTGGTCCG	CGTGGGTTTG	TCCGCCGCCG
45	1001	TCAGCCTCGT	GTTCAATCAT	TTTGTAACCG	ACTTCCCGAT	GGACATTTCC
	1051	GCCATGTCCG	TCATCGGCGC	GGTCGCCTGT	TCGACCGGAA	TCGGCATCGC
	1101	GTTCCGGCTT	ATGCTGGCCA	ATAAAGCAGC	CAAATCAAT	CCGATAGATG
	1151	CATTGGCGCA	GGATTGA			

This encodes a protein having amino acid sequence (SEQ ID NO: 536):

50 1 MSVOAVLAHK MRSLLTMLGI IIGIASVVSV VALNGSQKK ILEDISSIGT

51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT  
 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK  
 151 DKL FADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM  
 201 HQITGESHTN SITV KIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI  
 5 251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA  
 301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS  
 351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD\*

10 ORF134a (SEQ ID NO: 536) and ORF134-1 (SEQ ID NO: 534) show 100.0% identity in 388 aa overlap:

orf134a.pep	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG
orf134-1	MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG
orf134a.pep	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
15 orf134-1	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
orf134a.pep	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
orf134-1	RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
20 orf134a.pep	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITV KIKDNANTQVAEKGLTDLLKARHGTE
orf134-1	ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITV KIKDNANTQVAEKGLTDLLKARHGTE
orf134a.pep	DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
25 orf134-1	DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
orf134a.pep	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
orf134-1	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVAC
30 orf134a.pep	STGIGIAFGFMPANKAAKLNPIDALAQDX
orf134-1	STGIGIAFGFMPANKAAKLNPIDALAQDX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF134 (SEQ ID NO: 532) shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) (SEQ ID NO: 538) from *N. gonorrhoeae*:

35 orf134.pep	ARHGTE	DFFMNSDXIRQIVESTTGTMKLL	30
orf134ng	GESHTNSITV KIKDNANTRVAEKGLAELLKARHGTE	DFFMNSDSIRQMVESTTGTMKLL	264
orf134.pep	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG		90
40 orf134ng	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIGG		324

-398-

```

orf134.pep    LVGVGLSAAVSLVFNHFVTDPFMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 150
              |||||
orf134ng      LVGVGLSAAVSLVFNHFVTDPFMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 384

orf134.pep    LAQD    154
              ||||
5 orf134ng      LAQD    388

```

The complete length ORF134ng nucleotide sequence (SEQ ID NO: 537) is:

```

10      1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATCGTTTCGC TTCTGACCAT
      51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG
     101  GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG
     151  AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GGCGCAGCGG
     201  CAAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
     251  GCTACGTTGC CTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACC
     301  TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
     351  TTTTCGACGTG CGCGGGCTGA AGCTGGAAAC GGGGCGGCTG TTTGATGAGA
     401  ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA
     451  GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAACCA TTTGTTCAG
     501  GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAACGCTT
     551  TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
     601  CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
     651  AGACAATGCC AATACCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCTCA
     701  AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
     751  AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
     801  CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGTGTG ATGAACATTA
     851  TGCTGGTGTC CGTTACCGAG CGCACCAAG AAATCGGCAT ACGGATGGCA
     901  ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
     951  GTTAATCTGC ATCATCGGAG GCTTGTCG CGTAGGTTTG TCCGCCCGCG
    1001  TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTTCG
    1051  GCGGCATCCG TTATCGGGGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
    1101  GTTCGGCTTT ATGCCTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG
    1151  CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 538):

```

35      1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALNGSQKK ILEDISSMGT
      51  NTISIFPGRG FGDRRSBKIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
     101  YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
     151  DKLFDSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
     201  HQITGESHTN SITVLIKDNA NTRVAEKGLA ELLKARHGTE DFFMNNSDSI
     251  RQMVESTTGT MKLLISSIAL ISLVVGIGV MNIMLVSVTE RTKEIGIRMA
     301  IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
     351  AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

ORF134ng (SEQ ID NO: 538) and ORF134-1 (SEQ ID NO: 534) show 97.9% identity in 388 aa

45 overlap:

```

50      orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSMGTNTISIFPGRG
              |||||
      orf134-1      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG

      orf134ng      FGDRRSBKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
              |||||:|||||
      orf134-1      FGDRRSBKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

```

```

5  orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
    orf134-1     RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

    orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGTE
    orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE

    orf134ng      DFFMNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
    orf134-1     DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

10  orf134ng      IGARRGNILQQFLIEAVLICIGGLVGVGLSAAVSLVFNHFVTD FPMDISAASVIGAVAC
    orf134-1     IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTD FPMDISAMSVIGAVAC

    orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
15  orf134-1     STGIGIAFGFMPANKAAKLNPIDALAQDX

```

ORF134ng (SEQ ID NO: 538) also shows homology to an *E.coli* ABC transporter (SEQ ID NO: 1145):

```

20  sp|P75831|YBJZ_ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ )gi5
    (AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length = 648
    Score = 297 bits (753), Expect = 6e-80
    Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

    Query: 1  MSVQAVLAHKMRSLTMLXXXXXXXXXXXXXXXXXGNGSQKILEDISSMGNTNTISIFPGRG 60
    M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
25  Sbjct: 260 MAWRALAANKMRTLTLMLGIIIGIASVVSIVVVGDAAKQMVLA DIRSIGTNTIDVYPGKD 319

    Query: 61  FGDRRSQGIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
    FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
    Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

    Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
    G+ G F++ + AQVVVD N + +LF +D +G+ IL P VIGV ++
30  Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLD SNTRRQLFPHKADVVGEVILVGNMPARVIGVAEE 439

    Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
    ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
    Sbjct: 440 KQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFD SAEAEQQLTRLLSLRHGK 499

    Query: 240 EDFFMNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 299
    +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
35  Sbjct: 500 KDFFTWNMDGV LKTVKTTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559

    Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTD FPMDISAASVIGAVA 359
    A+GAR ++LQQFLIE F+ + + S +++ A
40  Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAF 619

    Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
    CST GI FG++PA AA+L+P+DALA++
    Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 65

5 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 539):

```

1  ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCCTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCTTCCTT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
10  201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCGCGATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCTTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
15  451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
501 TATGACCGTC GTTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

20 This corresponds to the amino acid sequence (SEQ ID NO: 540; ORF135):

```

1  ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIISAV
25  201 F*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 541):

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
30  101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
35  301 ACCCTGAGTT ACACCTCGTC GATTTTTTGG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCT TTTACACGCA GGCGGTGCTG CTCCTTGTT
401 TTGCCGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTGGT TTGGGCGACG
40  601 CTGACCGGCT GGCACACCT GTCCCTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGGCGTG TCCGCCTGA TTGCCAACT GTCGATGACG CGCGCCTACA
701 AAGTCGGCGA CAAATTCACG GTTGCCCTCG TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTCTG GCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
45  851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 542; ORF135-1):

-401-

5  
1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS  
51 TVALGAAAVL RRDxFRTPHW KNHLNRSMTVG TGAMLLLFYA VTHLPLATGV  
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSQE  
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVWAT  
201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR  
301 \*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 (SEQ ID NO: 540) shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) (SEQ ID NO: 544) from strain A of *N. meningitidis*:

15  
orf135.pep  
orf135a  
20  
orf135.pep  
orf135a  
25  
orf135.pep  
orf135a  
30  
orf135a  
35

```

                                10      20      30
                                GTGAMLLLFYAVTILPLATGVTLSTSSIF
                                |||||
                                50      60      70      80      90      100
STVALGAAAVLRRDTRTPHWKNHLNRSMTVGAMLLLFYAVTHLPLATGVTLSTSSIF

                                40      50      60      70      80      90
LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSQETAALAGLAGGAMSGWAYLK
LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSQETAALAGLAGGAMSGWAYLK
110      120      130      140      150      160

                                100      110      120      130      140      150
VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSCIGVSALIAQLSM
VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSCIGVSALIAQLSM
170      180      190      200      210      220

                                160      170      180      190      200
TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVFX
TRAYKVGDKFTVASLSYMTVVFSALSAAFFLAELFWQEILGMCIIILSGILSSIRPTAF
230      240      250      260      270      280

KQRLQSLFRQRX
290      300

```

The complete length ORF135a nucleotide sequence (SEQ ID NO: 543) is:

40  
45

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCG
51  GCGGCGCTGC TTTACCATT TGAACGTATT GATTAAAGAG GCATCGGCAG
101 AATTGTCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGCGCA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCT TTTACACGCA GGCGGTGCTG CTCCTTGCTT
401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTTCGCAG CCGTCAGGAA
451 ACGCGCGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG

```

551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG  
 601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCTGTG  
 651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA  
 701 AAGTCGGCGA CAAATTCACG GTTGCCTCGC TTTCCTATAT GACCGTCGTT  
 751 TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA  
 801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA  
 851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA  
 901 TAA

10 This encodes a protein having amino acid sequence (SEQ ID NO: 544):

1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLFS  
 51 TVALGAAAVL RRDTRTPHW KNHLNRSMTG TGAMLLLFYA VTHLPLATGV  
 101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE  
 151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT  
 201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
 251 FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR  
 301 \*

20 ORF135a (SEQ ID NO: 544) and ORF135-1 (SEQ ID NO: 542) show 99.3% identity in 300 aa overlap:

orf135a.pep	MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
orf135-1	MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
orf135a.pep	RRDTRTPHWKNHLNRSMTGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
orf135-1	RRDXFRTPHWKNHLNRSMTGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE
orf135a.pep	RISVYTQAVLLLGAFAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
orf135-1	RISVYTQAVLLLGAFAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
orf135a.pep	WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSMTRAYKVGDKFT
orf135-1	WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSCIGVSALIAQLSMTRAYKVGDKFT
orf135a.pep	VASLSYMTVVFSAALSAAFFLAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
orf135-1	VASLSYMTVVFSAALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 (SEQ ID NO: 540) shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) (SEQ ID NO: 546) from *N.gonorrhoeae*:

orf135.pep	GTGAMLLLFYAVTXLPLATGVTLSYTSSIF	30
orf135ng	STVTLGAAAVLRRDTRTPHWKNHLNRSMTGTGAMLLLFYAVTHLPLTTGVTLSYTSSIF	335



orf135.pep	LAVFSFLILKERISVYTQAVLLLLGFAGVLLLNPSFRSQETAALAGLAGGAMSGWAYLK	90
orf135ng	LAVFSFLILKERISVYTQAVLLLLGFAGVLLLNPSFRSQEPAALAGLAGGAMSGWAYLK	395
orf135.pep	VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM	150
orf135ng	VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM	455
orf135.pep	TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF	201
orf135ng	TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF	506

An ORF135ng nucleotide sequence (SEQ ID NO: 545) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 546):

1	MPSEKAFRRH	LRTASFQGLH	LHHFHQKVGK	CGIIGFGIHI	FPTLLPAAQG
51	ILDIQLGLFR	IDFAALAVYR	RTQVDFIHTV	IDGIASDQAF	SEVVQILRRRL
101	NLGHFTDTHL	IAQARRFIAD	FGNIRPMRRG	EAKTFCRCFR	FDGIDGIHGD
151	FRQCGHINRL	APGKDCRNGK	RDKVFFHTRH	YNQVCLEKTN	CSARKIKFRH
201	QKQAKTHSTS	LAARFTIRPS	LSQRPFMDTA	KKDILGSGWM	LVAACFTVM
251	NVLIKEASAK	FALGSGELVF	WRMLFSTVTL	GAAAVLRRDT	FRTPHWK NHL
301	NRSMVGTGAM	LLLFYAVTHL	PLTTGVTLST	TSSIFLAVFS	FLILKERISV
351	YTQAVLLLG	AGVLLLNPS	FRSQEPAAL	AGLAGGAMSG	WAYLKVRELS
401	LAGEPGWRVV	FYLSATGVAM	SSVWATLTGW	HTLSFPSAVY	LSGIGVSALI
451	AQLSMTRAYK	VGDKFTVASL	SYMTVVFSAL	SAAFFLGEEL	FWQEILGMC
501	IISAAF*				

Further work revealed the following gonococcal sequence (SEQ ID NO: 547):

1	ATGGATACCG	CAAAAAAGA	CATTTTAGGA	TCGGGCTGGA	TGCTGGTGCG
51	GGCGGCCTGC	TTCACCGTTA	TGAACGTATT	GATTAAAGAG	GCATCGGC
101	AATTTGCCCT	CGGCAGCGGC	GAATTGGTCT	TTTGGCGCAT	GCTGTTTCA
151	ACCGTTACGC	TCGGTGCTGC	CGCCGTATTG	CGGCGCGACA	CCTTCCGCAC
201	GCCCCATTGG	AAAAACCACT	TAAACCGCAG	TATGGTCGGG	ACGGGGGCGA
251	TGCTGCTGCT	GTTTTACGCG	GTAACGCATC	TGCCTTTGAC	AACCGGCGTT
301	ACCCTGAGTT	ACACCTCGTC	GATTTTTtg	GCGGTATTTT	CCTTCCTGAT
351	TTTGAAAGAA	CGGATTTCCG	TTTACACGCA	GGCGGTGCTG	CTCCTTGGTT
401	TTGCCGGCGT	GGTATTGCTG	CTTAATCCCT	CGTTCCGCAG	CGGTCAGGAA
451	CCGGCGGCAC	TCGCCGGGCT	GGCGGGCGGC	GCGATGTCCG	GCTGGGCGTA
501	TTTGAAAGTG	CGCGAACTGT	CTTTGGCGGG	CGAACCCGGC	TGGCGCGTCG
551	TGTTTTACCT	TTCCGCAACC	GGCGTGGCGA	TGTCGTCggt	ttgggcgacg
601	Ctgaccggct	ggCACAcccT	GTCCTTTTcca	tggcgagttt	ATCtgtCGGG
651	CATCGGCGTG	tccgcgCtgA	TTGCCCAaCT	GtcgatgAcg	cGCGcctaca
701	aaGTCGGCGA	CAAATTCACG	GTTGCCTCGC	tttcctaTAt	gaccgtcGTC
751	TTTTCCGCCC	TGTCTGCCGC	ATTTTTTCTg	ggcgaagagc	ttttCtggCA
801	GGAAATACTC	GGTATGTGCA	TCATTAtccT	CAGCGGCATT	TTGAGCAGCA
851	TCCGCCCCAT	TGCCTTCAAA	CAGCGGCTGC	AAGCCCTCTT	CCGCCAAAGA
901	TAA				

This corresponds to the amino acid sequence (SEQ ID NO: 548; ORF135ng-1):

1	MDTAKKDILG	SGWMLVAAAC	FTVMNVLKE	ASAKFALGSG	ELVFWRMLFS
51	TVTLGAAAVL	RRDTFRTPHW	KNHLNRSMVG	TGAMLLLFYA	VTHLPLTTGV
101	TLSTSSIFL	AVFSFLILKE	RISVYTQAVL	LLGFAGVLL	LNPSFRSQE
151	PAALAGLAGG	AMSGWAYLKV	RELSLAGEPG	WRVVFYLSAT	GVAMSSVWAT
201	LTGWHTLSFP	SAVYLSGIGV	SALIAQLSMT	RAYKVGDKFT	VASLSYMTVV

251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR  
 301 \*

ORF135ng-1 (SEQ ID NO: 548) and ORF135-1 (SEQ ID NO: 542) show 97.0% identity in 300 aa overlap:

```

  orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRMLFSTVTLGAAAVL
  orf135-1       MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMLFSTVALGAAAVL

  orf135ng-1.pep RRDTFRTPHWKNHLNRSVMVGTGAMLLLFYAVTHLPLTTGVTLSTSSIFLAVFSFLILKE
  orf135-1       RRDXFRTPHWKNHLNRSVMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE

  orf135ng-1.pep RISVYTQAVLLLGFAGVVLLLNPFSRSGQEPAAAGLAGGAMSGWAYLKVRELSLAGEPG
  orf135-1       RISVYTQAVLLLGFAGVVLLLNPFSRSGQETAAGLAGGAMSGWAYLKVRELSLAGEPG

  orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
  orf135-1       WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT

  orf135ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR
  orf135-1       VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 66

The following DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 549):

```

  1  ATGAAGCGGC GTATAGCCGT CTTCGTCCTG TTCCCGCAGA TAATCCGAGT
  51  TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
  101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT
  151 CTGCCCCGGA TCGCCGAAAT CGATTCCCA TCGGCATCG TGTTCGGTGC
  201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
  251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
  301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAGTTCG CCsGGTTCAT
  351 TGTTTCAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC
  401 CACATATGTT CGCAAATTTT GCCGTCTTCG CCGTCTTGGA AAAAAGGGAC
  451 TTTGACCATG GCAAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCCCAAA
  501 AAAGcTCGCG CCAAAAATAT TTGAATGTTT TACGGGCGCG TTCGTCCGCA
  551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC
  601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTTGCA CCTTACTGCG
  651 GCTTTCTgcC kTCGGCATCC GATTTCGATT TGAAAAGTTC mmrwyATTCG
  701 GAATAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 550; ORF136):

5

1	MKRRIAVFVL	FPQIRVLGQ	LLPKIVNTVP	AHRMLFQIFG	MMMMFIHQOY
51	LPGIAEIDSP	CGIVFGALLF	RHLPAHCLYG	KAAVGDAVAH	EHVPADVVR
101	NANAFALFDI	GQFAXFIVQH	TVNIKTVKIN	IVDPHMFANF	AVFAVLEKRD
151	FDHGKIQQGN	NAAAFPKKLA	PKIFECFTGA	FVGTVYRFVC	LFYIINDGIA
201	HHSAPORVRY	LFAFPYCGFLP	SASDSDLKSS	XXSE*	

Further work revealed the complete nucleotide sequence (SEQ ID NO: 551):

	1	ATGATGAAGC	GGCGTATAGC	CGTCTTCGTC	CTGTTCCCGC	AGATAATCCG
10	51	AGTTTTGGGA	CAACTGTTGC	CGAAAAATCGT	CAATACAGTT	CCGGCACATC
	101	GGATGCTCTT	CCAGATTTTC	GGGATGTTCT	TTTTCTTCAT	ACACCAGCAA
	151	TATCTGCCCG	GGATCGCCGA	AATCGATTCC	CCATGCGGCA	TCGTGTTCCG
	201	TGCGCTCCTC	TTCCGTATC	TGCCCGCGCA	TTGCCCTGTAT	GGTAAAGCCG
	251	CCGTAAGGGA	TGCGTTGCA	CACGAACATC	CAGTCGCTGA	TGTCGTCAAC
15	301	CGGAACGCAA	ACGCTTTCGC	CTTGTTGCAC	ATTGGTCAGT	TCGCCGGGTT
	351	CATTGTTTCA	CACACCGTAA	ATATAAAGAC	CGTCAAAATA	AATATCGTCG
	401	ATCCACATAT	GTTTCGCAAAT	TTCGCCGTCT	TCGCCGTCTT	GGAAAAAAGG
	451	GACTTTGACC	ATGGCAAAAT	CCAAGGCGGA	AATAATGCGG	CGGCGTTCCC
	501	AAAAAAGCTC	GCGCCAAAAA	TATTTGAATG	TTTTACGGGC	GCGTTCGTCG
20	551	GCACGGTTTA	CCGGTTCGTC	TGCCTGTTCT	ACATAATAAA	TGACGGAATC
	601	GCCCATCATT	CTGCTCCTCA	ACGTGTACGG	TATCTGTTTG	CACCTTACTG
	651	CGGCTTTCTG	CCTTCGGCAT	CCGATTCGGA	TTTGAAAAGT	TCCAAATATT
	701	CGGAATAG				

This corresponds to the amino acid sequence (SEQ ID NO: 552; ORF136-1):

25 1 MMKRRIAVFV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ  
51 YLPGIAEIDS PCGIVFGALL FRHLPACLY GKAAVGDAVA HEHPVADVVN  
101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHFAN FAVFAVLEKR  
151 DFDHGKIQQG NNAAAFPKKL APKIFECFTG AFVGTVYRVF CLFYIINDGI  
30 201 AHHSAPQVRV YLFAPYCGFL PSASDSDLKS SKYSE\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 (SEQ ID NO: 550) shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) (SEQ ID NO: 554) from strain A of *N. meningitidis*:

35

	10	20	30	40	50	59
orf136.pep	MKRRIAVFVLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLAEIDS					
	:     :					
orf136a	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGLAEIDS					
	10	20	30	40	50	60

40

	60	70	80	90	100	110	119
orf136.pep	PCGVIFGALLFRHLPAHCLYGKAAGDAVAHEHPADVNNRANAFALFDIGQFAXFIVQ						
	:         :           :						
orf136a	PCGVIFGTLLFRHXSTHCLYGKAAGNAVAHEHPADVNNRANAFALFDIGQFAGFIVQ						
	70	80	90	100	110	120	

45

	120	130	140	150	160	170	179
orf136.pep	HTVNIKTVKINIVDPHFANFAVFVLEKRFDFHGKIQQGNNAAAFPPKKLAPKIFECFTG						
	::   :                               : :   :   :   :   :   :   :						

```

orf136a      HAINVKTVKINIVDPHMFANFAXFAVLEKRALTMAKSKXXXMRRRSQKSSRQKYLNLRA
              130      140      150      160      170      180

orf136.pep   180      190      200      210      220      230
              AFVGTVYR VFCLFYIINDGIAHH---SAPQVRVRYLFAPYCGFLPSASDSLKSSXXSEX
              :  ||: | :  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf136a      R---SPARFTGLSACSTXXMTESPIISAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX
              190      200      210      220      230

```

The complete length ORF136a nucleotide sequence (SEQ ID NO: 553) is:

10	1	ATGATGAAGC	GGCGTATAGC	CGTCTTCGTC	CTGCTCATGC	AGAAAAATCCG
	51	GATTTTGGGA	CAACTGTTGC	CGAAAAATCGT	CAATACAGTT	CCGGCACATC
	101	GGATGCTCTT	CCAGATNTTC	GGGATGTTCT	TTTTCTTCAT	ACACCAGCAA
	151	TACCTGCCCG	GGATCGCCGA	AATCGATTCC	CCATGCGGCA	TCGTGTTCCG
	201	TACGCTCCTC	TTCCGTCATC	NGTCCACGCA	TTGCCTGTAT	GGTAAAGCCG
15	251	CCGTAGGGAA	TGCCGTTGCA	CACGAACATC	CAGTCGCTGA	TGTCGTCAAC
	301	CGGAACGCAA	ACGCTTTTCG	CTTGTTTCGAC	ATTGTCAGT	TCGCCGGGTT
	351	CATTGTTTCA	CACGCCATAA	ATGTAAAGAC	CGTCAAAATA	AATATCGTCG
	401	ATCCACATAT	GTTCGCAAAT	TTCGCCNTCT	TCGCCGTCTT	GGAAAAAAGG
	451	GCTTTGACCA	TGGCAAAATC	TAAGGNGNNA	NNGATGCGGC	GGCGTTCCCA
20	501	AAAAAGCTCG	CGCCAAAAAT	ATTTGAATGT	TTTGCGGGCG	CGTTCGCCCG
	551	CACGGTTTAC	CGGTTTGTCT	GCCTGTTCTA	CATAATAAAT	GACGGAATCG
	601	CCCATCATAT	CTGCTCCTCA	ACGTGTACGG	TATCTGTTTG	CACCTTACTG
	651	CGCCTTTCTG	CCTTCGGCAT	CCGATTCCGA	TTTGAAAAGT	TCCAAATATT
	701	CGGAATAG				

25 This encodes a protein having amino acid sequence (SEQ ID NO: 554):

```

1  MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNAVA HEHPVADVVN
101 RNANAFALFD IGQFAGFIVQ HAINVKTVKI NIVDPHMFAN FAXFAVLEKR
151 ALTMAKSXXX XMRRRSQKSS RQKYLNVLR RSPARFTGLS ACST**MTES
201 PIISAPORVR YLFAPYCGFL PSADSDLKS SKYSE*

```

ORF136a (SEQ ID NO: 554) and ORF136-1 (SEQ ID NO: 552) show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
35	orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFFIHQOYLPGIAEIDS					
	orf136-1	MMKRRIAVFVLFPPQIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFFIHQOYLPGIAEIDS					
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf136a.pep	PCGIVFGTLLFRHXSTHCLYGKAAVGNVAHEHPVADVNNRANAFALFDIGQFAGFIVQ					
	orf136-1	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNNRANAFALFDIGQFAGFIVQ					
		70	80	90	100	110	120
		130	140	150	160	170	180
45	orf136a.pep	HAINVKTVKINIVDPHMFANFAXFAVLEKRALTKMAKSXXXMRRRSQKSSRQKYLNVLRA					
	orf136-1	HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPPKKLAPKIFECFTG					
		130	140	150	160	170	180
		190	200	210	220	230	

```

orf136a.pep  R---SPARFTGLSACSTXXMTESPIISAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX
              : ||: | : :: | ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf136-1     AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX
              190      200      210      220      230

```

## 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 (SEQ ID NO: 550) shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) (SEQ ID NO: 556) from *N.gonorrhoeae*:

```

10 orf136.pep  MKRRIAVFVLFPPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHHQYLPGLAEIDS 59
    ||||| : | ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf136ng      MMKRRIVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGLAEIDS 60

    orf136.pep  PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQFAXFIVQ 119
    | ||||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf136ng      PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ 120

15 orf136.pep  HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKIFECFTG 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : |||||
orf136ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAAFPKKLAPKVFEFTG 180

    orf136.pep  AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSLKSSXXSE 234
    ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 orf136ng      AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSLKSSKYSE 235

```

The complete length ORF136ng nucleotide sequence (SEQ ID NO: 555) is:

```

1  ATGATGAAGC GCGTATAGC CGTCTTCGTC CTGCTCATGC AGAAATCCG
51  GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
25 101 GGATGCTCTT CCAAATTTTC GGGATGTTCT TTTTCTTCAT ACACCGGCAA
    151 TACCTGCCCC GGATCGCCGA AATCGATTCC CCAGCGGTA TCGTGTTCCG
    201 TACGCTCCTC TTCCGTCATC TGTCGCGCA TGCCTGTAC GGTAAAGCCG
    251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTGCGCAAC
    301 CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT CCGCCGGGTT
    351 CATTGTTCAG CACACCGTAA ATATAAGAC CGTCAAAATA AATATCGTCG
30 401 ATCCACATAT GTTCGCAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAGG
    451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
    501 AAAAAAGCTC GCGCCAAAAG TATTTGAATG TTTACGGGC GCGTTCGCCG
    551 GCACGGTTTA CCGGTTTCGT TGCCTGTTCT ACATAATAAA TGACGGAATC
    601 GCCCATCATA CTGCTCCTCA ACGTGTACGG TATCTGTTG CACCTTACCG
35 651 CGGTTTTCTA CCTCCGCAT CCGATTCGGA TTTGAAAAGT TCCAATATT
    701 CGGAATAG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 556):

```

40 1  MMKRRIVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFIHRQ
    51  YLPGLAEIDS PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEHPVADVAN
    101  RNANAFALFD IGQSAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
    151  DFDHGKIQGG NNAAAFPKKL APKVFECFTG AFAGTVYRFV CLFYIINDGI
    201  AHTAPQVRV YLFAPYRGFL PPASDSLKS SKYSE*

```

ORF136ng (SEQ ID NO: 556) and ORF136-1 (SEQ ID NO: 552) show 93.6% identity in 235 aa overlap:

```

5      orf136ng      MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
      orf136-1      MMKRRIAVFVLFPPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS

      orf136ng      PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQSAGFIVQ
      orf136-1      PCGIVFGALLFRHLPACHLYGKAAVGDAVAHEHPADVNRNANAFALFDIGQFAGFIVQ

10     orf136ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAPFKKLAPKVFECEFTG
      orf136-1      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDFDHGKIQGGNNAAPFKKLAPKIFECFTG

      orf136ng      AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSDLKSSKYSEX
      orf136-1      AFVGTVYRFVCLFYIINDGIAHSAPQVRVRYLFAPYCGFLPSASDSDLKSSKYSEX
15

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 67

20 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 557):

```

      1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
      51 CGCCGCCGCG TTGCTTGCCG CC.TGCGGAC GCGGGGAAAT AATGCTGTCC
101  GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
25  151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
      201 GAAAGAAAAC GGTATTCTTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
      251 CGATTGTCGG CAACCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
      301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
      351 CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAATTAC ATCAACCGAA
30  401 AACTCCGCGG CATGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCC..

```

This corresponds to the amino acid sequence (SEQ ID NO: 558; ORF137):

```

      1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
      51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLE
35  101 LEAEILGKTD LVDLTLSTNG FIKGAKLQNY INRKLGRMQI QQFPKIFAA..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 559):

```

      1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
      51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGGAAAT AATGCTGTCC
40  101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
      151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
      201 GAAAGAAAAC GGTATTCTTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT

```

5  
10  
15

```

251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCC GA CCGCCTCGAA
301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
351 CACCACTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC
501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG
551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
601 CCCGTCAGTG CCGCCCGGCG GCAGGGGGCG AATTTCTGTA TTGCCGTCGA
651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG
751 GGCAGGCGG ATGTGGTTAT CAAACGCAG GTTTTGGATT TGGGTGCAGT
801 CGCGGATTG GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG
851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
901 TGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 560; ORF137-1):

20

```

1  MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
51  GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE
101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV PQPVIIGRHT YVDGGLSQPV
201 PVSAARRQGA NFVIAVDISA RPKKNISQGF FSYLDQTLNV MSVSALQNEL
251 GQADVVIKPK VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
301 *

```

25 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF137 (SEQ ID NO: 558) shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) (SEQ ID NO: 562) from strain A of *N. meningitidis*:

30

```

          10      20      30      40      50      60
orf137.pep  MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf137a     MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
          10      20      30      40      50      60

          70      80      90      100     110     120
orf137.pep  VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTSTNG
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf137a     VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTSTSG
          70      80      90      100     110     120

          130     140     149
orf137.pep  FIKGAKLQNYINRKLGRMQIQFPIKFAA
          |||:|||||:|:|||||:|||||
orf137a     FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
          130     140     150     160     170     180

```

45 The complete length ORF137a nucleotide sequence (SEQ ID NO: 561) is:

```

1  ATGGAATAA TGGTAACGTT TTCAAAATC AGACCGCTTT TGGCAATCGC
51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGGAAAT AATGCTGCCC

```

5  
10  
15

```

101  GCAAGCCGGT  GCAAACCGCC  AAACCCGCCG  CAGTGGTCGG  TTTGGCACTC
151  GGTGGCGGCG  CATCTAAAGG  ATTTGCCCAT  GTAGGTATTA  TTAAGGTTTT
201  GAAAGAAAAC  GGTATTCCTG  TGAAGGTGGT  TACCGGCACA  TCGGCAGGTT
251  CGATAGTCGG  CAGCCTTTTT  GCATCGGGTA  TGTGCCCCGA  CCGCCTCGAA
301  TTGGAAGCCG  AAATTTTAGG  TAAAACCGAT  TTGGTCGATT  TAACCTTGTC
351  CACCAGTGGT  TTTATCAAAG  GCGAAAAGCT  GCAAAATTAC  ATCAACCGAA
401  AAGTCGGCGG  CAGGCGGATT  CAGCAGTTTC  CCATCAAATT  TGCCGCCGTT
451  GCTACTGATT  TTGAAACCGG  CAAGGCCGTC  GCTTTCAATC  AAGGGAATGC
501  CGGGCAGGCT  GTGCGCGCTT  CCGCCGCCAT  TCCCAATGTG  TTCCAACCCG
551  TTATCATCGG  CAGGCATACA  TATGTTGACG  GCGGTCTGTC  GCAGCCCGTG
601  CCCGTCAGTG  CCGCCCGGCG  GCANGNNNG  NATNTCGTGA  TTGCCGTCGA
651  TATTTCCGCC  CGTCCGAGCA  AAAACATCAG  CCAAGGCTTC  TTCTCTATC
701  TCGATCAGAC  GCTGAACGTA  ATGAGCGTTT  CCGCGTTGCA  AAATGAGTTG
751  GGGCAGGCGG  ATGTGGTTAT  CAAACCGCAG  GTTTTGGATT  TGGGTGCAGT
801  CGGCGGATTC  GATCAGAAAA  AACGCGCCAT  CCGGTTGGGT  GAGGAGGCAG
851  CACGTGCCGC  ATTGCCTGAA  ATCAAACGCA  AACTGGCGGC  ATACCGTTAT
901  TGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 562):

20  
25

```

1  MENMVTFSKI  RPLLAIAAAA  LLAACGTAGN  NAARKPVQTA  KPAAVVGLAL
51  GGGASKGFAH  VGIKVLKEN  GIPVKVVTGT  SAGSIVGSLF  ASGMSPDRLE
101 LEAEILGKTD  LVDLTLSTG  FIKGEKLQNY  INRKVGRRRI  QQPFIKFAAV
151 ATDFETGKAV  AFNQGNAGQA  VRASAAIPNV  FQPVIIGRHT  YVDGGLSQPV
201 PVSAAARRXX  XXVIAVDISA  RPSKNISQGF  FSYLDQTLNV  MSVSALQNEL
251 GQADVVIKPQ  VLDLGAVGGF  DQKKRAIRLG  EEAARAALPE  IKRKLAAARY
301 *

```

ORF137a (SEQ ID NO: 562) and ORF137-1 (SEQ ID NO: 560) show 97.3% identity in 300 aa overlap:

30  
35  
40

```

orf137a.pep  MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
orf137-1      MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH

orf137a.pep  VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTG
orf137-1      VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTG

orf137a.pep  FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
orf137-1      FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV

orf137a.pep  FQPVIIGRHTYVDGGLSQPVVPVSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV
orf137-1      FQPVIIGRHTYVDGGLSQPVVPVSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV

orf137a.pep  MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY
orf137-1      MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

```

45 Homology with a predicted ORF from *N. gonorrhoeae*



ORF137 (SEQ ID NO: 558) shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) (SEQ ID NO: 564) from *N.gonorrhoeae*:

```

5      orf137.pep  MENMVTFSKIRPLLAIAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH  60
      orf137ng    MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH  60

      orf137.pep  VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMS PDRLELEAEILGKTDLVDLTLSTNG  120
      orf137ng    IGIVKVLKENGIPVKVVTGTSAGSIVGSL LASGMS PDRLELEAEILGKTDLVDLTLSTSG  120

10     orf137.pep  FIKGAKLQNYINRKL RGMQIQFPPIKFAA  149
      orf137ng    FIKGEKLQNYINRKVGGRQIQFPPIKFAAVATDFETGKAVAFNQG NAGQAVRASAAIPNV  180

```

The complete length ORF137ng nucleotide sequence (SEQ ID NO: 563) is:

```

15      1  ATGGAATA  TGGTAACGTT  TTCAAAATC  AGATCATTTT  TGGCAATCGC
      51  CGCCGCCGCG  TTGCTTGCCG  CCTGCGGTAC  GCGGGGAAAC  AATGCCGCCC
      101  GCAAGCCGGT  GCAAACCGCC  AAACCCGCCG  CAGTGGTCGC  TTTGGCACTC
      151  GGTGGCGGCG  CATCTAAAGG  ATTTGCCCAT  ATAGGAATG  TTAAGGTTT
      201  GAAAGAAAAC  GGTATTCCTG  TGAAGGTGGT  TACCGGCACA  TCGGCAGGTT
      251  CGATAGTCGG  CAGCCTTTTG  GCATCGGGTA  TGTCGCCCGA  CCGCCTCGAA
20     301  TTGGAAGCCG  AGATTTTAGG  TAAACCGAT  TTAGTCGATT  TAACCTTGTC
      351  CACCACTGGT  TTTATCAAAG  GCGAAAAGCT  GCAAAATTAC  ATCAACCGAA
      401  AAGTCGGCGG  CAGGCAGATT  CAGCAGTTTC  CCATCAAATT  TGCCGCCGTT
      451  GCCACTGATT  TTGAAACCGG  CAAGGCCGTC  GCTTTCAATC  AAGGGAATGC
      501  CGGGCAGGCG  GTTCGTGCTT  CCGCCGCCAT  TCCCAATGTG  TTCCAGCCAG
25     551  TCATCATCGG  CAGGCACAAA  TATGTTGACG  GCGGTCTGTC  GCAGCCCGTG
      601  CCCGTCAGTG  CCGCTCGGCG  GCAGGGGGCG  AATTTCTGTA  TTGCCGTCGA
      651  TATTTCGCA  CGTCCGAGCA  AAAATGTCGG  TCAAGGTTTC  TTCTCTTATC
      701  TCGATCAGAC  GCTGAACGTG  ATGAGCGTTT  CCGTGTGCA  AAACGAGTTG
      751  gggcAGGCGG  ATGTGGTTAT  CAAACCGCag  gtTTTGGATT  TGGGTGCAGT
30     801  CGGCGGATTC  GATCAGAAAA  AGCGCGCCAT  CCGGTTGGGC  GAGGAGGCAG
      851  CACGTGCCGC  ATTGCCTGAA  ATCAAACGCA  AACTGGCGGC  ATACCGTTAT
      901  TGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 564):

```

35      1  MENMVTFSKI  RSFLAIAAAA  LLAACGTAGN  NAARKPVQTA  KPAAVVALAL
      51  GGGASKGFAH  IGIVKVLKEN  GIPVKVVTGT  SAGSIVGSL  ASGMS PDRLE
      101  LEAEILGKTD  LVDLTLSTSG  FIKGEKLQNY  INRKVGGRQI  QQFPPIKFAAV
      151  ATDFETGKAV  AFNQG NAGQA  VRASAAIPNV  FQPVIIGRHK  YVDGGLSQPV
      201  PVSAARRQGA  NFVIAVDISA  RPSKNVGQGF  FSYLDQTLNV  MSVSVLQNEL
40     251  GQADVVIKPQ  VLDLGA VGGF  DQKKRAIRLG  EEAARAALPE  IKRKLAA YRY
      301  *

```

ORF137ng (SEQ ID NO: 564) and ORF137-1 (SEQ ID NO: 560) show 96.0% identity in 300 aa overlap:

```

45     orf137ng    MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH
      orf137-1    MENMVTFSKIRPLLAIAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH

```

```

orf137ng      IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLDLTLSTSG
orf137-1      VGIKVLKENGIPVKVVTGTSAGSIVGSLLFASGMSPDRLELEAEILGKTDLDLTLSTSG

5  orf137ng      FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
orf137-1      FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV

orf137ng      FQPVIIGRHKYVDGGLSQPVPVSAARRQGANFVIAVDISARPSKNVGQGFYSYLDQTLNV
orf137-1      FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGFYSYLDQTLNV

10 orf137ng      MSVSVLQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY
orf137        MSVSALQNELGQADVVIKPQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 68

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 565):

```

20      1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
      51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT
     101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
     151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CGGGTTTGAA
     201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TCGGGAACG GCAAAGGCG
25      251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
     301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGACAA
     351 ACACGAAGG CTGCTATTC.

```

This corresponds to the amino acid sequence (SEQ ID NO: 566; ORF138):

```

30      1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGLHAFYLL
      51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET
     101 MFKAVHGWEH VQQALDKHEG LLF

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 567):

```

35      1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
      51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT
     101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
     151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
     201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TCGGGAACG GCAAAGGCG
40      251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
     301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGACAA
     351 ACACGAAGG CTGCTATTC TACGCCGCA CATCGGCAGC TACGATTGG
     401 GCGGACGCTA CATCAGCCAG CAGCTCCGT TCCCGCTGAC CGCCATGTAC

```

451 AAACCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT  
 501 TCGCGGCAAA GGAAAAACCG CGCCTACCAG CATACAAGGG GTCAAAACAAA  
 551 TCATCAAAGC CCTGCGTTCG GCGGAAGCAA CCATCGTCTT GCCCGACCAC  
 601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG  
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG  
 701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT  
 751 TTCGATTGTC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC  
 801 CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT  
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This corresponds to the amino acid sequence (SEQ ID NO: 568; ORF138-1):

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGH LAFYLL  
 51 KEDRARIVAN MRQAGLNPDP KTVKAVFAET AKGGLELAPA FFRKPEDIET  
 101 MFKAVHGWEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY  
 151 KPPKIKAIK IMQAGRVGRK GKTAPTSIQG VKQIIKALRS GEATIVLPDH  
 201 VPSPQEGGEG VWVDFFGKPA YTMTLAAKLA HVKGVKTLFF CCERLPGGQG  
 251 FDLHIRPVQG ELNGDKAHDA AVFNRNAEYW IRRFPTQYLF MYNRYKMP\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 (SEQ ID NO: 566) shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) (SEQ ID NO: 570) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf138.pep	MFRLQFRLFP	PLRTAMHILL	TALLKCLSL	LPLSCLHTLGN	RLGH LAFYLL	KEDRARIVAX	
orf138a	MFRLQFRLFP	PLRTAMHILL	TALLKCLSL	LPLSCLHTLGN	RLGH LAFYLL	KEDRARIVAN	
		10	20	30	40	50	60
orf138.pep	MRQAGLNPDP	KTVKAVFAET	AKGGLELAPA	FFRKPEDIET	MFKAVHGWEH	VQALDKHEG	
orf138a	MRQAGLNPDP	KTVKAVFAET	AKGGLELAPA	FFRKPEDIET	MFKAVHGWEH	VQALDKHEG	
		70	80	90	100	110	120
orf138.pep	LLF						
orf138a	LLFITPHIGSYDLGGYISQ	QLPFPLTAMY	KPPKIKAIKIMQAGRVGRK	GKTAPTSIQG			
		130	140	150	160	170	180

The complete length ORF138a nucleotide sequence (SEQ ID NO: 569) is:

1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGCAGG CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGTCAGG CAGGCATGAA  
 201 TCCCCAGCCC AAAACGGTCA AAGCCGTTT TCGCGAAACG GCAAAGGCG  
 251 GTTTGGAAC TGCCTCCGCG TTTTTCAGAA AACCAGGAAGA CATAGAAACA  
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGGG CTGCTATTCA TCACGCCGCA CATCGGCAGC TACGATTGG

5

10

401	GCGGACGCTA	CATCAGCCAG	CAGCTTCCGT	TCCCGCTGAC	CGCCATGTAC
451	AAACCGCCGA	AAATCAAAGC	GATAGACAAA	ATCATGCAGG	CGGGCAGGGT
501	TCGCGGCAAA	GGAAAAACCG	CGCTTACCAG	CATACAAGGG	GTCAAACAAA
551	TCATCAAAGC	CCTGCGTTCT	GGCGAAGCAA	CCATCGTCCT	GCCCGACCAC
601	GTCCCCTCCC	CTCAAGAAGG	CGGGGAAGGC	GTATGGGTGG	ATTTCTTCGG
651	CAAACCTGCC	TATACCATGA	CGCTGGCGGC	AAAATTGGCA	CACGTCAAAG
701	GCGTGA AAAA	CTGTTT TTT	TGCTGCGA AC	GCCTGCCTGG	CGGACAAGGT
751	TTCGATT TGC	ACATCCGCC C	CGTCCAAGGG	GAATTGAACG	GCGACAAAAGC
801	CCATGATGCC	GCCGTGTTCA	ACCGCAATGC	CGAATATTGG	ATACGCCGTT
851	TTCCGACGCA	GTATCTGTTT	ATGTACAACC	GCTACAAAAT	GCCGTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 570):

```

1  MFRLQFRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTLGN  RLGLHAFYLL
51 KEDRARIVAN  MRQAGLNPD  P  KTVKAVFAET  AKGGLELAPA  FFRKPEDIET
101 MFKAVHGWHEH  VQQALDKHEG  LLFITPHIGS  YDLGGYISQ  QLPFPFLTAMY
151 KPPKIKAIKD  IMQAGRVRG  K  GKTAPTSIQG  VKQIIKALRS  GEATIVLPDM
201 VPSPEQEGEG  VVWVDFGKPA  Y  YTMTLAAKLA  HVKGKTLFF  CCELRPGQG
251 FDLHIRPVQG  ELNGDKAHDA  AVFNRNAEYW  IRRFPTQYLF  MYNRYKMP*

```

20 ORF138a (SEQ ID NO: 570) and ORF138-1 (SEQ ID NO: 568) show 99.7% identity over a 298aa overlap:

[illegible]

### Homology with a predicted ORF from *N. gonorrhoeae*

ORF138 (SEQ ID NO: 566) shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) (SEQ ID NO: 572) from *N.gonorrhoeae*:

[illegible]

orf138.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDMETMFKAVHGWVQVQALDKHEG 120  
 orf138ng MRQAGLNPDQTQTVKAVFAETAKCGLELAPAFFKKPEDMETMFKAVHGWVQVQALDKGEG 120  
 5 orf138.pep LLF 123  
 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG 180

The complete length ORF138ng nucleotide sequence (SEQ ID NO: 571) is:

1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 10 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCACACG CAGACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAATGCG  
 251 GTTTGGAAC TGGCCCCGCG TTTTCAAAA AACCAGGAAG CATCGAAACA  
 15 301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA  
 351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGCG  
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC  
 451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT  
 501 GCGCGGCAAA GGCAAAACcg cgcccaccgg catACAAGGG GTCAAAACAA  
 20 551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CCATcATCCT GCCCGACCAC  
 601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA  
 651 ACCTGCATAc acCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG  
 701 TGAACCCCT GTTTTCTGCG TGCGAACGCC TGCCCGACGG ACAAGGCTTC  
 751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAGCCCCA  
 25 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTT  
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAAACGCC GTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 572):

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL SLSCLHTLGN RLGHAFYLL  
 30 51 KEDRARIVAN MRQAGLNPDQTQTVKAVFAET AKCGLELAPA FFKKPEDIET  
 101 MFKAVHGWVH VQVQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY  
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQG VKQIIKALRA GEATIIILPDH  
 201 VPSPQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDQGQF  
 35 251 VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP\*

ORF138ng (SEQ ID NO: 572) and ORF138-1 (SEQ ID NO: 568) show 94.3% identity over 299aa overlap:

orf138-1.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHAFYLLKEDRARIVAN  
 40 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLSLSCLHTLGNRLGHAFYLLKEDRARIVAN  
 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDMETMFKAVHGWVQVQALDKHEG  
 orf138ng MRQAGLNPDQTQTVKAVFAETAKCGLELAPAFFKKPEDMETMFKAVHGWVQVQALDKGEG  
 45 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTISIQG  
 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG  
 orf138-1.pep VKQIIKALRSGEATIVLPDHVPSPQEGGGVWVDFFGKPAYTMTLAAKLAHVKGKVTLFF  
 orf138ng VKQIIKALRAGEATIIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGKVTLFF

gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253  
Score = 80.8 bits (196), Expect = 9e-15  
Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)

Query: 160 KIMQAGVRVGKGTAPTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGVWADFFGKPA 219  
 ++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A  
 Subject: 151 ELLRKORVOLGNKVAASTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVPFFATQA 208

Query: 220 YTMTLAAKLAHVKGVKTLFFCCERLPDQGQF 250  
T + +F RLPDG G+  
Spict: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

ORF138-1 (SEQ ID NO: 568) (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 14B). These experiments confirm that ORF138-1 (SEQ ID NO: 568) is a surface-exposed protein, and that it is a useful immunogen.

30 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 573):

35

```

      1  . . GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
     51  GCATGCGGTG TGGAATACTT TGCCTTCTC GCGGCGGCG GTGTATGCGG
    101  CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CGGCGCGCG GTCGGCGTGG
    151  ATGCGCGGGC TGATGTTTA GCCGTTATG GTGTCGCCG TTTGTGTTTC
    201  GGCGGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTC TTGCCGTTGC
    251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGG AAAAGATGTT
    301  TTATCAGCCT GGGATGCACT GCCGCGGAT TACGGCAGG CGGCGCGCGG
    351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
```

401 TGAAACCGGC GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG  
 451 GGCGAATTTG CGGCGACATT GTTCTGTCG CGTCCGGAAT GGCAGACGCT  
 501 GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG  
 551 CGCGGGCGAT GGTGCTG..

5

This corresponds to the amino acid sequence (SEQ ID NO: 574; ORF139):

1 .AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGVV YAAPARRSAW  
 51 MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV  
 101 LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV  
 151 GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

10

Further work revealed the complete nucleotide sequence (SEQ ID NO: 575):

1 ATGGATGGAC GGC GTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC  
 51 GGCTTTTGTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT  
 101 ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA  
 151 CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT  
 201 GCCTTTGGGC GTGCCTGTCG CGTGGGTGCT GGCGGGGCTG GCGTTTCCGG  
 251 GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCACG  
 301 TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT  
 351 GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTAC GGCAATGTGT  
 401 TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTGTGCAA  
 451 GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG  
 501 GCGGCGGTTT TGGGACATTG AAATGCCCCG TTTGCGCCCC TGGCTTGCCG  
 551 GCGGCGTGTG CCTGTCTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG  
 601 CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA  
 651 GTTGCTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC  
 701 TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC  
 751 AGGCGCGCGG TTTCGGATAA GGCGGTTTCC CTTGTGATGC CGTCGCCGCC  
 801 GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTCGGCG GCGGTGTTGT  
 851 CTGTGTGCTG CCTGTTTCTT TTGTTGGCAA TTGTTGTGAA AGCGTGGTCG  
 901 GCCGCGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT  
 951 GTGGAATACT TTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT  
 1001 TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTCCGGCTG GATGCGCGGG  
 1051 CTGATGTTTT TGCCGTTTAT GGTGTCGCCG GTTGTGTTT CGGCGGCGGT  
 1101 GCTGTGCTT TATCCGCAGT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA  
 1151 TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC  
 1201 TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTGGGTGTC  
 1251 AAACGCTTT CAGACGGCAT GCCGCATCAC GTCCCCCTC TTGAAACCGG  
 1301 CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT  
 1351 GCGGCGACAT TGTTCCTGTC GCGTCCGGA TGGCAGACGC TGACGACTTT  
 1401 GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA  
 1451 TGGTGTGAC ATTGCTGTTG GCGGCGTTTC CGCTGGGTAT TTTCTGCTG  
 1501 TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

45 This corresponds to the amino acid sequence (SEQ ID NO: 576; ORF139-1):

1 MDGRRWVWVG AFALLPSAFL AVMVVAPLWA VAAVDGLAWR AVLSDAYMLK  
 51 RLAWTVFQAA ATCVLVPLG VPAWVLARL AFPGRALVLR LLMLPFVMP  
 101 LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFVQ  
 151 VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA  
 201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG  
 251 RRAVSDKAVS PVMPSPQSV GEYVLLAFAA AVLSVCCFLP LLAIIVVKAWS  
 301 AGESWRVLME SETWQAVWNT LRFSAAAVYA AAVLGVVYAA AARRSAWMRG  
 351 LMFLPFMVSP VCVSAGVLLL YPQWTASLPL LLAMYALLAY PFVAKDVLSA  
 401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF

50

451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL  
501 LDGGEGGKQT ETL\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF139 (SEQ ID NO: 574) shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) (SEQ ID NO: 578) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF139a nucleotide sequence (SEQ ID NO: 577) is:

30	1	ATGGATGGAC	GGCGTTGGGC	GGTATGGGGT	GCTTTTGCCC	TGCTGCCTTC
	51	GGCTTTTGTG	GCGGCAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
	151	CGTTTGGCGT	GGACGGTATT	TCAGGCAGCG	GCAACCTGTG	TGCTGGTGCT
35	201	GCCTTTGGGC	GTGCCTGTCT	CGTGGGTGCT	GGCGCGGCTG	GCGTTTCCGG
	251	GGCGGGCTTT	GGTGCTGCGC	CTGCTGATGC	TGCCTTTTGT	GATGCCACG
	301	TTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTCGGGGCGG	ACGGCCTGTN
	351	GTGGCGCGCG	TGGCAGGATA	CGCCGATATC	GTTGTTGTAT	GGCAATGTGT
40	401	TTTTTNACCT	TCCTGTGTTG	GTCAGGCGCG	CATATCAGGG	GTTTGTGCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACNG	ACATTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCCG	TTTGCGCCCG	TGGCTTGCCG
	551	GCGGCGTGTG	CCTTGTCTTC	CTGTATTGTT	TTTCGGGGTT	CGGGCTGGCA
45	601	TTGCTGCTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTGGTCAAT	TTCGAACTCG	ATATGCCGGT	TGCTTCGGTG	CTNGTGTGGC
	701	TGGTGTNNGG	GGTAACNCGC	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTCGGATAA	GGCNGTTTCC	CCTGTGATGC	CGTCGCCCGC
	801	GCACTCGGTC	GGGGAATATG	TGCTNCTGCC	GTTTGCGGCG	CGGGTGTNGT
	851	CTGTGTGCTG	CCTGTTTCNT	TTGTTTGCAA	TTGTTGTGAA	AGCGTGCTCG
	901	GCCGGCGAAT	CGTGGCGTGT	GTTAATGGAA	AGTGAACCGT	GGCAGGCGGT



5	951	GTGGAATACT	NTGCGCTTCT	CGGCGGCGGC	GGTGTATGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGTCGCGGTG	GATGCGCGGG
	1051	CTGATGTTT	TGCCGTTTAT	GGTGTGCGCG	GTTTGTGTTT	CGGCGGGCGT
	1101	GCTGCTGCTT	NATCCGCAGT	GGACGGCTTC	GTGCGCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCAGCC
10	1201	TGNGATGCAC	TGCCGCGGGA	TTACGGCAGG	GCGGCGGCGG	TTTTGGGTGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGCATCAC	GTTCCCCCTC	TTGAAAACGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CAACCTGCGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTNTGTC	GCGTCNCGAG	TGGCAGACGC	TGACGACTTT
	1401	GATTTATGCC	TATNTGGGAG	GCGCGGGTGA	NGATAATTAC	GCGCGGGCGA
	1451	TGTTGCTGAC	ATTGCTGTTG	GCGGCGTTTC	CGCTGGGTAT	NTTCTTGCTG
	1501	TTGGACGGCG	GCGAAGCGGG	AAAACGGACG	GAACCGTTAT	AA

This encodes a protein having amino acid sequence (SEQ ID NO: 578):

15	1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPGL	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGVLAL	FGADGLXWRG	WQDTPYLLLY	GNVFFXLPVL	VRAAYQGFVQ
	151	VPAARLQTAX	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	<u>LYCFSGFGLA</u>
	201	<u>LLGG</u> SRYAT	VEVEIYQLVM	FELDMAVASV	<u>LVWL</u> VXGVTA	<u>AAGLL</u> YAWFG
20	251	RRAVSDKA VS	PVMPSPQSV	GEYVLLAFAA	AVXSVCCFLX	LLAIVVKAWS
	301	AGESWRVLME	SETWQAVWNT	XRFSAAVYA	<u>AAVL</u> GVVYAA	AARRSAWMRG
	351	LMFLPFMVSP	VCVSAGVLLL	XPQWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	XDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
	451	AATLFXSRXE	WQTLTTLIYA	YXGRAGXDNY	ARAMVLTLLL	<u>AAFALGXFL</u> L
25	501	LDGGEGGKRT	ETL*			

ORF139a (SEQ ID NO: 578) and ORF139-1 (SEQ ID NO: 576) show 96.5% homology over a 514aa overlap:

30	orf139a.pep	MDGRRWAVWGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA      :     :
	orf139-1	MDGRRWVVWGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA      :     :
	orf139a.pep	ATCVLVLPPLGVPVAVWLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLAALFGADGLXWRG 
	orf139-1	ATCVLVLPPLGVPVAVWLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLAALFGADGLLWRG 
35	orf139a.pep	WQDTPYLLLYGNVFFXLPVLVRAAYQGFVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP 
	orf139-1	RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP 
	orf139a.pep	WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVLXGVTA 
	orf139-1	WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA 
40	orf139a.pep	AAGLLYAWFGRRRAVSDKAVSPVMPSPPPQSVGEYVLLAFAAAVXSVCCLFXLLAIIVKAW 
	orf139-1	AAGLLYAWFGRRRAVSDKAVSPVMPSPPPQSVGEYVLLAFAAAVLSVCCLFPLLAIIVKAW 
	orf139a.pep	AGESWRVLMESSETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP 
	orf139-1	AGESWRVLMESSETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP 
45	orf139a.pep	VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLSAXDALPPDYGRAAGLGANGF 
	orf139-1	VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF 

-420-

```

orf139a.pep  QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNY
orf139-1      QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY

orf139a.pep  ARAMVLTLLLAALGXFLLLDGGEGGKRTETLX
orf139-1      ARAMVLTLLLAALGXFLLLDGGEGGKRTETLX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 (SEQ ID NO: 574) shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) (SEQ ID NO: 580) from *N.gonorrhoeae*:

```

10  orf139.pep                               AWSAGESWRVLMESETWHAVWNTLRFSA   30
                                     ||||| |||||:|||||
orf139ng  QSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWSAGESRRVLMESETWQAVWNTLRFSA   327

orf139.pep  VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLLAM   90
           |:||||| |||||:|||||:||||| ||||| ||||| ||||| ||||| ||||| |||||
15  orf139ng  VFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSPVCVSAGVLLLYPQWTASLPLLLAM   387

orf139.pep  LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV   150
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf139ng  LAYPFVAKDVLSAWDALPPDYGRAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV   447

orf139.pep  GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL                     189
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20  orf139ng  GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLDN   507
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

The complete length ORF139ng nucleotide sequence (SEQ ID NO: 579) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 580):

```

25      1  MDGRCAVVRG  AFSLLPSAFL  AVMVVAPLWA  VAAYDGLAWR  AVLSDAYMLK
      51  RLAWTVFQAA  ATCVLVPLG  VPVAWVLARL  AFPGRALVLR  LLMLPFVMP
101  101  LVAGVGV LAL  FGADGLWRG  RQDTPYLLLY  GNVFFNLPVL  VRAAYQGFAQ
151  151  VPAARLQTAR  TLGAGAWRPF  WDIEMPVLRP  WLAGGVCLVF  LYCFSGFGLA
201  201  LLLGGSRYAT  VEVEIYQLVM  FELDMAGASA  LVWLVLGVTA  AAGLLYAWFG
30  251  RRAVSDKAVS  PVMPSPQSV  GEYVLLAFSV  AVLSVCCLFP  LSAIVVKAWS
301  301  AGESRRV LME  SETWQAVWNT  LRFSAAVFA  AAVLGVVYAA  AARRLVWMRG
351  351  LVFLPFMVSP  VCVSAGVLLL  YPGWTASLPL  LLAMYALLAY  PFVAKDVLSA
401  401  WDALPPDYGR  AAAGLGANGF  QTACRITFPL  LKPALRRGLT  LAAATCVGEF
451  451  AATLFLSRPE  WQTLTTLIYA  YLGRAGEDNY  ARAMVLTLLL  SAFAVCIFLL
35  501  LDNGEGGKRT  ETL*

```

Further work revealed a variant gonococcal DNA sequence (SEQ ID NO: 581):

```

40      1  ATGGATGGAC  GGTGTTGGGC  GGTACGGGGT  GCTTTTCC   TGCTGCCTTC
      51  GGCTTTT   TGG  GCGGTAATGG  TCGTTGCGCC  TTTGTGGGCG  GTGGCGGCGT
101  101  ATGACGGTT  GCGGTGGCGC  GCGGTGCTGT  CGGATGCCTA  TATGCTCAAA
151  151  CGTTTGGCGT  GGACGGTGTT  TCAGGCGGCG  GCAACCTGTG  TGCTGGTGCT
201  201  GCCTTTGGGC  GTGCCTGTCT  CGTGGGTGCT  GGCGCGGCTG  GCGTTCCTCG
251  251  GCGCGGCTT  GGTGCTGCGC  CTGCTGATGC  TGCCGTTTGT  GATGCCACG
301  301  CTGGTGGCGG  GCGTGGGCGT  GCTGGCTCTG  TTCGGGCGG  ACGGGCTGTT

```

	351	GTGGCGCGGC	CGGCAGGATA	CGCCGTATCT	GTGTGTGTAC	GGCAATGTGT
	401	TTTTCAACCT	GCCCGTGTG	GTCAGGCGG	CGTATCAGGG	GTTTGCTCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGGCGTG
5	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGCGCCCG	TGGCTTGCCG
	551	GCGGCGTGTG	CCTTGTCTTC	CTGTATTGTT	TTTCGGGGTT	CGGGCTGGCA
	601	TTGCTGTTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTGGTTATG	TTCGAACTCG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGGTGTGGG	GGTAACGGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
10	751	AGGCGCGCGG	TTTCGGATAA	GGCGGTTTCC	CCCCTGATGC	CGTCGCCGCC
	801	GCAATCGGTG	GGGGAATATG	TATTGCTGGC	ATTTTCGGTG	GCGGTGTTGT
	851	CCGTGTGCTG	CCTGTTTCCT	TTGTCGGCAA	TTGTTGTGAA	AGCGTGGTCG
	901	GCCGGCGAAT	CGCGCGGTGT	GTTAATGGAA	AGTGAAACGT	GGCAGGCGAGT
	951	GTGGAATACT	tGCGCTTTT	CGGCGGCGGC	GGTGTTTGCG	GCGGCGGTTT
15	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGCTGGTGTG	GATGCGCGGA
	1051	CTGGTGTTTT	TACCGTTTAT	GGTGTGCGCG	GTTTGTGTTT	CGGCGGGCGT
	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCGGCG
	1201	TGGGATGCAC	TGCGCCGGA	TTACGGCAGG	GCGGCGGCAGT	GTTTGGGCGC
20	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CGACGTGTGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCCCTGTC	GCGTCCGGAA	TGGCAGACGT	TGACGACTTT
	1401	GATTTATGCC	TATTTGGGGC	GTGCGGGTGA	GGACAATTAT	GCGCGGGCAA
	1451	TGGTGTTGAC	ATTGCTGTTG	TCGGCATTTG	CGGTGTGCAT	TTTCTGCTG
25	1501	TTGGACAACG	GCGAAGGCGg	aaaACGGACG	GAACGTTTAT	AA

This corresponds to the amino acid sequence (SEQ ID NO: 582; ORF139ng-1):

	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
30	101	<u>LVAGVGV LAL</u>	FGADGLLWRG	RQDTPYLLLY	GNVFFNLVPL	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRRF	WDIEMPVLRP	<u>WLAGGVCLVF</u>	<u>LYCFSGFGLA</u>
	201	<u>LLLGG</u> SRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCFLP	LSAIVVKAWS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	<u>AAVLGVVYAA</u>	AARRLVWMRG
35	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
	451	AATLFLSRPE	WQTLTLLIYA	YLGRAGEDNY	ARAMVLTLLL	<u>SAFAVCIFLL</u>
	501	LDNGEKKRT	ETL*			

ORF139ng-1 (SEQ ID NO: 582) and ORF139-1 (SEQ ID NO: 576) show 95.9% identity over  
40 513aa overlap:

[illegible]

	orf139ng	AAGLLYAWFGRRVSDKAVSPVMPSPQSVGEYVLLAFSAVLSCCLFPLSAIVVKAWS 
	orf139-1	AAGLLYAWFGRRVSDKAVSPVMPSPQSVGEYVLLAFSAVLSCCLFPLLAIVVKAWS 
5	orf139ng	AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP 
	orf139	AGESWRVLMESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP 
	orf139ng	VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF 
	orf139-1	VCVSAGVLLLYPQWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF 
10	orf139ng	QTACRITFLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY 
	orf139-1	QTACRITFLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY 
	orf139ng	ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL 
15	orf139-1	ARAMVLTLLLSAALGIFLLLDGEGGKQTETL 

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or  
20 diagnostics, or for raising antibodies.

### Example 70

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 583):

25

```

1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC
101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
151 GGTGTGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
251 AACGTTTGGT C...
```

30 This corresponds to the amino acid sequence (SEQ ID NO: 584; ORF140):

```

1  MDGWTQTLQA QTLGISAIAA IILILILIVR FRIHALTLV IVSLLTALAT
51  GLPTGSIVKD ILVKNFGGTL GGVALLVGLG AMLERLV...
```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 585):

35

```

1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
151 GGTGTGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
40 251 GACGTTTGGT CGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG
301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
```

5  
10  
15  
20

```

401  TGCCCATCGT  GTTCGCCACC  GCACGGCGCA  TGAAACAGGA  CGTACTGCCC
451  TTCGCGCTTG  CCTCCATCGG  CGCATTTTCC  GTCATGCACG  TCTTCCTGCC
501  GCCCCATCCG  GGCCCGATTG  CCGCTTCCGA  ATTTTACGGC  GCGAACATCG
551  GCCAAGTTTT  GATTTTGGGT  CTGCCGACCG  CCTTCATCAC  ATGGTATTTT
601  AGCGGCTATA  TGCTCGGCAA  AGTGTGTTGG  CGCACCATCC  ATGTTCCCGT
651  TCCCGAACTG  CTCAGCGGCG  GCACGCAAGA  CAACGACCTG  CCGAAAGAAC
701  CTGCCAAAGC  AGGAACGGTC  GTCGCCATCA  TGCTGATTCC  CATGCTGCTG
751  ATTTTCTCTG  ATACCGGCGT  ATCGGCCCTC  ATCAGCGAAA  AACTCGTAAG
801  TCGCGACGAA  ACCTGGGTTC  AGACGGCAAA  AATAATCGGT  TCGACACCGA
851  TCGCCCTTCT  GATTTCGGTA  TTGGTTCGCAC  TGTTTGTCTT  GGGACGCAAA
901  CGCGGCGAAA  GCGGCAGCGC  GTTGGAAAAA  ACCGTGGACG  GCGCACTCGC
951  CCCCCTCTGT  TCCGTGATTC  TGATTACCGG  CGCGGGCGGT  ATGTTCCGGC
1001 GCGTTTTGCG  CGCTTCCGGC  ATCGGCAAGG  CACTCGCCGA  CAGCATGGCG
1051 GATTTGGGCA  TTCCCGTCCT  TTTGGGCTGT  TTCCTTGTCG  CCTTGGCACT
1101 GCGTATCGCG  CAAGTTCGG  CAACCGTCGC  CCTGACCACC  GCCGCCGCGC
1151 TGATGCTCC  TGCCGTTCGC  GCCGCCGGCT  TTACCGACTG  GCAGCTCGCC
1201 TGTATCGTAT  TGGCAACGGC  GGCAGTTTCG  GTCGGTTGCA  GCCACTTCAA
1251 CGACTCCGGC  TTCTGGCTGG  TCGGCCGTCT  CTTGGACATG  GACGTACCGA
1301 CCACGTGAA  AACCTGGACG  GTCAACCAAA  CCCTCATCGC  ACTCATCGGC
1351 TTTGCCTTGT  CCGCACTGCT  GTTCGCCATC  GTCTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 586; ORF140-1):

25  
30

```

1  MDGWTQTLA QTLGISA AAA IILILILIVK FRIHALLTLV IVSLLTALAT
51  GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVL P
151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIQVLIILG LPTAFITWYF
201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
251 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK
301 RGESGSALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSMA
351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
401 CIVLATAAGS VGCSHFND SG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
451 FALSALLFAI V*

```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 (SEQ ID NO: 584) shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) (SEQ ID NO: 588) from strain A of *N. meningitidis*:

40  
45

```

              10      20      30      40      50      60
orf140.pep  MDGWTQTLAQTLLGISAAAIIILILILIVRFRIHALLTLVIVSLLTALATGLPTGSIVKD
              |||
orf140a     MDGWTQTLAQTLLGISAAAIIILILILIVKFRHALLTLVIVSLLTALATGLPTGSIVND
              10      20      30      40      50      60

              70      80
orf140.pep  ILVKNFGGTLGGVALLVGLGAMLERLV
              :
orf140a     VLVKNFGGTLGGVALLVGLGAMLERLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
              70      80      90      100     110     120

```

The complete length ORF140a nucleotide sequence (SEQ ID NO: 587) is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCTGT TGGGCATTTT  
 51 GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 151 GGT TTGCCCA CAGGCAGCAT TGTC AACGAC GTACTGGTCA AAAACTTCGG  
 5 201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GGCGGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT GTTCGCGACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 10 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC  
 501 GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT  
 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 15 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCTCTGA ATACCGCGCT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TGCGGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA  
 901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 20 951 CCGCGTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC  
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTCT CCTTGCACT  
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACT GCCGCGCGCG  
 1151 TGATGGCTCC TGCGGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC  
 25 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTCT GTCGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCTT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

30 This encodes a protein having amino acid sequence (SEQ ID NO: 588):

1 MDGWTQTLISA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP  
 151 FALASIGAFS VMHVFLPPHP GPAAASEFYG ANIGQVLILG LPTAFITWYF  
 35 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 251 IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGESGSALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSMA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSHFND SG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG  
 40 451 FALSALLFAI V\*

ORF140a (SEQ ID NO: 588) and ORF140-1 (SEQ ID NO: 586) show 99.8% identity over a 461aa overlap:

45 orf140-1.pep MDGWTQTLISAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 orf140a MDGWTQTLISAQTLGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 orf140-1.pep ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 : ||||| 120  
 orf140a VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 50 orf140-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPAAASEFYG 180  
 ||||| 180  
 orf140a GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPAAASEFYG 810

	orf140-1.pep	ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV	240
	orf140a	ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV	240
5	orf140-1.pep	VAIMLIPMLLI FLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK	300
	orf140a	VAIMLIPMLLI FLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK	300
	orf140-1.pep	RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC	360
	orf140a	RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC	360
10	orf140-1.pep	FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGC SHFNDSG	420
	orf140a	FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGC SHFNDSG	420
	orf140-1.pep	FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV	461
15	orf140a	FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV	461

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 (SEQ ID NO: 584) shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) (SEQ ID NO: 590) from *N.gonorrhoeae*:

20	orf140.pep	MDGWTQTL SAQTLLGISAAAIILILILIVRFRIHALLTLVIVSLLTALATGLPTGSIVKD	60
	orf140ng	MDGRTQTL SAQTLLGISAAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND	60
	orf140.pep	ILVKNFGGTLGGVALLVGLGAMLERLV	87
25	orf140ng	VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF	120

The complete length ORF140ng nucleotide sequence (SEQ ID NO: 589) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 590):

	1	MDGRTQTL SA	QTLLGISAAA	IILILILIVK	FRIRALLTLV	IASLLTALAT
30	51	GLPTGSIVND	VLVKNFGGTL	GGVALLVGLG	AMLGRLVETS	GGAQSLADAL
	101	IRMFGEKRAP	FAPGVASLIF	GFPIFFDAGL	IVMLPIVFAT	ARRMKQDVLP
	151	FALASVGAFS	VMHVFLPPHP	GPIAASEFYG	ANIGQVLILG	LPTAFITWYF
	201	SGYMLGKVLG	RAIHVPVPEL	LSGGTQSDP	PKEPAKAGTV	VAVMLIPMLL
	251	IFLNTGVSAL	ISEKLVSAD E	TWVQTAKMIG	STPVALLISV	LAALLVLGRK
	301	RGESGSTLEK	TVDGALAPAC	SVILITGAGG	MFGGVLRASG	IGKALAD SMA
35	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDWQLA
	401	CIVLATAAGS	VGCSHFND SG	FWLVGR LSDM	DVPTTLKTWT	VNQT LIAFIG
	451	FALSALLFAI	V*			

Further work revealed a variant gonococcal DNA sequence (SEQ ID NO: 591):

40	1	ATGGACGGCC	GGACACAGAC	GCTGTCCGCG	CAAACCTTGT	TGGGCAT TTC
	51	GGCGGCGGCA	ATCATCCTCA	TTCTGATTTT	AATCGTCAAA	TTCCGCATCC
	101	GCGCGCTGCT	GACACTGGTC	ATCGCCAGCC	TGCTGACGGC	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	CGTCAACGAC	GTACTGGTCA	AAAAC TTCG

	201	CGGCACGCTC	GGCGGCGTGG	CGCTTCTGGT	CGGTCTGGGC	GCAATGCTCG
	251	GACGTTTGGT	AGAAACATCC	GGCGGCGCAC	AGTCGCTGGC	GGACGCGCTG
	301	ATCCGGATGT	TCGGCGAAAA	ACGCGCACCG	TTCGCTCCGG	GCCTGCGCTC
5	351	GCTGATTTTC	GGCTTCCCGA	TTTTCTTCGA	TGCCGGACTA	ATCGTCATGC
	401	TGCCCATCGT	ATTCGCCACC	GCACGGCGCA	TGAAACAGGA	CGTACTGCCC
	451	TTCCGCGTTG	CCTCCGTCGG	CGCATTTTCC	GCATGACACG	TCTTCCTGCC
	501	TCCCCATCCG	GGCCCCGATTG	CCGCTTCCGA	ATTTTACGGC	GCGAACATCG
	551	GCCAGGTTTT	GATTTTGGGT	CTGCCGACCG	CCTTCATCAC	ATGGTATTTT
10	601	AGCGGCTATA	TGCTCGGCAA	AGTGTTGGGG	CGCGCCATCC	ATGTTCCCGT
	651	TCCCGAATCG	CTCAGCGCGC	GCACGCAAGA	CAGCGACCCG	CCGAAAAGAA
	701	CTGCCAAAGC	AGGAACGGTC	GTCCGCCGTA	TGCTGATTTT	CATGCTGCTG
	751	ATTTTCTCTA	ATACCGCGGT	ATCAGCCCTC	ATCAGCGAAA	AACTCGTAAG
	801	TGCGGACGAA	ACTTGGGTTC	AGACGGCAAA	AATGATCGGT	TCGACACCTG
15	851	TCGCCCTTCT	GATTTCCGTA	TTGGCCGCAC	TGTTGGTCTT	GGGACGCAAA
	901	CGCGGCGAAA	GCGGCAGCAC	GTTGGAAAAA	ACCGTGGACG	GCGCACTCGC
	951	CCCCGCGTGT	TCCGTGATTC	TGATTACCGG	CGCGGGCGGT	ATGTTCCGGC
	1001	GCGTTTTGCG	CGCTTCCGGC	ATCGGCAAGG	CACCTCGCCG	CAGCATGGCG
	1051	GATTTGGGCA	TTCCCGTCCT	TTTGGGTGTC	TTTCTGTCTG	CCTTGGCACT
20	1101	GCGTATCGCG	CAAGGTTTCG	CAACCGTCGC	CCTGACCACA	GCCGCCGCGC
	1151	TGATGGCTCC	TGCCGTTGCC	GCCGCCGGCT	TTACCGACTG	GCAGCTCGCC
	1201	TGTATCGTAT	TGGCAACGGC	GGCAGGTTTC	GTCGGTTGCA	GCCACTTCAA
	1251	CGACTCCGCG	TTCTGGCTGG	TCGGCCGCCT	CTTGATATAT	GACGTACCGA
	1301	CCACGCTGAA	AACCTGGACG	GTCAACCAAA	CCCTCATCGC	ATTTCATCGG
25	1351	TTTGCCTTGT	CCGCACTGCT	GTTTGCCATC	GTCTGA	

This corresponds to the amino acid sequence (SEQ ID NO: 592; ORF140ng-1):

	1	MDGRTQTL	SA QTL	LGISAAA	IILILILIVK	FRIRALLTLV	IASLLTALAT
	51	GLPTGSIVND	VLVKNF	GGTL	GGVALLVGLG	AMLGRLVETS	GGAQSLADAL
30	101	IRMFG	KEKRAP	FAPGVASLIF	GFPIFFDAGL	IVMLPIVFAT	ARRMKQDVLP
	151	<u>FALASVGA</u>	<u>FS VMHVF</u>	LPPHP	GPIAASEFYG	ANIGQVLIIG	LPTAFITWYF
	201	SGYMLGKVLG	RAIHVPVPEL	LSGGTQSDP	PKEPAKAGTV	<u>VAVMLIPMLL</u>	
	251	<u>IFLNTGVS</u>	<u>AL ISEKLV</u>	<u>SADE</u>	<u>TWVQTAKMIG</u>	<u>STPVALLISV</u>	<u>LAALLVLGRK</u>
	301	RGESGSTLEK	TVDGALAPAC	<u>SVILITGAGG</u>	<u>MFGGVL</u>	<u>LRAS</u>	IGKALADWSMA
35	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDWQLA	
	401	CIVLATAAGS	VGCSHFND	SG FWLVGRLLDM	DVPTTLTKWT	VNQTLIAFIG	
	451	FALSALLFAI	V*				

ORF140ng-1 (SEQ ID NO: 592) and ORF140-1 (SEQ ID NO: 586) show 96.3% identity over 461aa overlap:

40	orf140ng-1.pep	MDGRTQTL <del>SAQ</del> TLLGISAAAILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND
	orf140-1	MDGWTQTL <del>SAQ</del> TLLGISAAAILILILIVKFR <del>I</del> HALLTLVIVSLLTALATGLPTGSIVND
45	orf140ng-1.pep	VLVKNFGGTLGGVALLVGLGAMLGRLVETSSGAQSLADALIRMFGEKRAPFAPGVASLIF
	orf140-1	ILVKNFGGTLGGVALLVGLGAMLGRLVETSSGAQSLADALIRMFGEKRAPFALGVASLIF
	orf140ng-1.pep	GFPIFFDAGLIVMLPIVFATARRMQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG
	orf140-1	GFPIFFDAGLIVMLPIVFATARRMQDVLPPFALASIGAFSVMHVFLPPHPGPIAASEFYG
50	orf140ng-1.pep	ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSSGGTQDS <del>DP</del> PK <del>E</del> PAKAGTV
	orf140-1	ANIGQVLILGLPTAFITWYFSGYMLGKVLGR <del>T</del> IHVPVPELLSSGGTQDNDLPKEPAKAGTV



orf140ng-1.pep VAVMLIPMLLIFLNTGVSA LISEKLVSADETWWQTAKMIGSTPVALLISVLAALLVLGRK  
 orf140-1 VAIMLIPMLLIFLNTGVSA LISEKLVSADETWWQTAKIIGSTPIALLISVLVALFVLGRK  
 5 orf140ng-1.pep RGE SGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC  
 orf140-1 RGE SGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC  
 orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFND SG  
 orf140-1 FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFND SG  
 10 orf140ng-1.pep FWLVGRLLDMDVPTTLKTWTNVNQT LIAFIGFALSALLFAIV  
 orf140-1 FWLVGRLLDMDVPTTLKTWTNVNQT LIALIGFALSALLFAIV

Furthermore, ORF140ng-1 (SEQ ID NO: 592) is homologous to an *E.coli* protein (SEQ ID NO:  
 15 1148):

gi|882633 (U29579) ORF\_o454 [Escherichia coli] )gi|1789097 (AE000358) o454;  
 This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa  
 protein GNTP\_BACLI SW: P46832 [Escherichia coli] Length = 454  
 Score = 210 bits (529), Expect = 1e-53  
 Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)  
 20  
 Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147  
 E SGGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K  
 Sbjct: 80 EHS GGAESLAN YFSRKLGD KRTIAALTLAAFFLGIPVFFDVGFII LAP IYGF AKVAKIS 139  
 25  
 Query: 148 VLPFALASVGAFSVMHVFLPPHPGP IAA SEFYGANIGQV LILGLPTAFITWYFSGYMLGK 207  
 L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K  
 Sbjct: 140 PLKFG L P VAGIMLT VH VAVPPHPGPVAAAGLLHADIGWL TIIGIAIS-IPVG VVG YFAAK 198  
 Query: 208 VLGRAIHVPVPELL-----SGGTQSDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257  
 ++ + + E+L G T+ SD P A V ++++IP+ +I T  
 Sbjct: 199 IINKRQYAMSVEVLEQMQ LAPASEEGATKLS DKINPPGVA-LVTS LIVIPIAIIMAGT-- 255  
 30  
 Query: 258 SALISEKLVSADETWWQTAKMIGSTPXXXXXXXXXXXXXGRKRGE SGSTLEKTVDGALA 317  
 +S L+ + T ++IGS +RG S + AL  
 Sbjct: 256 ---VSATLMPPSHPLLGLTLQ LIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312  
 Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377  
 A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS  
 35 Sbjct: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMIDLPLLPAAFIISLALRASQGS--AT 370  
 Query: 378 XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFND SGFWLVGRLLDMDVPTTLK 437  
 G Q + LA G +G SH ND SGFW+V + L + V LK  
 Sbjct: 371 VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430  
 Query: 438 TWTVNQT LIAFIGFALSALLFAIV 461  
 TWTV T++ F GF ++ ++A++  
 40 Sbjct: 431 TWTVLTTILGFTGFLITWCVWAVI 454

Based on this analysis, including the identification of the presence of a putative leader sequence  
 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 71

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 593):

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
      51  TTTGCTGTCTG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
     151  AACTTTTTTGG GCAGACACCA CGGGCGCAC. GTCGTCTCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCAG TTGCCATTCTT CCTCAACCCC GCTGCCGCCG
     251  CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
     301  CGCGTGATTG CCGCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTGTC CCTGATGCTG CCCTTGCCCCG
     401  TACTGATGTT TTTCCGTCCG ..

```

15 This corresponds to the amino acid sequence (SEQ ID NO: 594; ORF141):

```

      1  ..DFGISPVYLW VAAAFKHLIS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
     51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAAFAAAGL VLHGYSLARR
    101  RVIAASFLLG TGWTLSLAA AYPAAFALML PLPVLMMFFRP ...

```

20 Further work revealed the complete nucleotide sequence (SEQ ID NO: 595):

```

      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
     51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
    101  TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
    151  TCGAAGCAC TGGCAGGACG CCCCACCCC TTGGTTGCCC ATCTGTTCGG
    201  TCAAACCGAT TTCGGCATA CCCCCGTGTA TCTTTGGGTT GCCGCCGCGT
    251  TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACTCATACTA TGCCGCACGC
    301  TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCCT GCGGCTTTGC
    351  CGGTTTCAAC TTTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
    401  TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT
    451  GCCCCGCCCT TTGCCCGCG CCGACTGGTG CTGCACGGTT ATTCTTTGGC
    501  TCGCCGCGCG GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGCACGC
    551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
    601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
    651  GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTATG ACCGTTTACC
    701  CGCTGCTCTT GGCAAAACG CAGCCCGCGC TGTTCCGCGA ATGGCTCGAC
    751  TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTTT AGACGGCATT
    801  CAGTTTGTTC TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
    851  TGCCCGTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
    901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
    951  CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCCGCCGC
   1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGGGCGC
   1051  GCGTTTGTCA ACTGGTTCCG CATTATGGCG TTCGGACTGT TTGCCGTGTT
   1101  CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
   1151  CCGAACGCGC CGCCTATTTC AGCCCGTATT ATGTTCTCTGA TATCGATCCC
   1201  ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT
   1251  TACCCGGAAC AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
   1301  GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
   1351  GACGCGGCGA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT
   1401  TTCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA

```



5  
10  
15  
20  
25  
30

```

151  GTCGAAGCAC TGGCAGGCAG CCCCACCCCT TTGGTTGCC ATCTGTTCCG
201  TCAAATCGAT TTCGGCATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
251  TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACCCGTATGA TGCCGCACGC
301  TTTGCCGGCG TGTTTTTTCGC CGTTGTCGGA CTGACTTCCT GCGGCTTTGC
351  CGGTTTCAAC TTTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTG
401  TCATCGGCTG TATCGGGCTG ATTCCGACCG TACACTTTCT CAACCCCGCT
451  GCCGCCGCCT TTGCCGCCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
501  TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGTGAGCGC
551  TGATGTCGTT GGCAGCAGCT TATCCGGCGG CATTGGCCCT GATGCTGCCC
601  CTGCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
651  GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
701  CGCTGCTCTT GGCAAAACG CAGCCCGCGC TGTTCGCGCA ATGGCTCGAC
751  GATCACGTTT TCGGTACGTT CGGCGGCGTG CCGCACATTG AGACGGCATT
801  CAGTTTGTTT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCTGCGC
851  TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TCGACCGAC
901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTTGG TGCTGCTTGC
951  CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTCCGCCGCG
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGACG CGGCGCGGCG
1051 GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGAAGTG TTGCCGTGTT
1101 CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CCGAACGCGC CGCCTATTTC AGCCCGTATT ATGTTCTGTA TATCGATCCC
1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGC TGTGGGCGAT
1251 TACCCGCAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGCT
1401 TTCCCGGAA TTAACACGGG AGCTTTCAGA CGGCATCGAG TGTATCGACA
1451 TAGGCGGCGG CGACCTACAC ACGCGGATTG TTTGGACGCA GTACGGCACA
1501 TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCCGCTT
1551 GCCCAAAAC GCGGATGCGC GCAAGGCTG GCAGACGGTC TGGCAGGGTG
1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
1651 GAAAATATAT TAAAAACAAC AGATTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 598):

35  
40  
45

```

1  MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
51  VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR
101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
151 AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD
251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCTRFLFSTD
301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYVVDIDP
401 IPMAVAVLFT PLWLWAIKTR NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT
501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKTG
551 ENILKTTD*

```

ORF141a (SEQ ID NO: 598) and ORF141-1 (SEQ ID NO: 596) show 98.2% identity in 553 aa overlap:

50

```

orf141a.pep  MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
              |||
orf141-1     MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP

orf141a.pep  LVAHLFGQIDFGIPPVYLWVAAFAKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
              |||
orf141-1     LVAHLFGQIDFGIPPVYLWVAAFAKHLLSPWAADSYDAARFAGVFFAVVGLTSCGFAGFN

```

-431-

	orf141a.pep	FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
	orf141-1	FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
5	orf141a.pep	GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPDLLAKT
	orf141-1	GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPDLLAKT
	orf141a.pep	QPALFAQWLDDHVFGTGGVVRHIQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD
	orf141-1	QPALFAQWLDYHVFGTGGVVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD
10	orf141a.pep	WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
	orf141-1	WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
15	orf141a.pep	FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAI TRK
	orf141-1	FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAI TRK
	orf141a.pep	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMELSPKRELSDGIE
	orf141-1	NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMELSPKRELSDGIE
20	orf141a.pep	CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVRLPQNADAPQGQWTVWQGARPRNKD
	orf141-1	CIGIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVLLPQNADAPQGQWTVWQGARPRNKD
	orf141a.pep	SKFALIRKTGENI
	orf141-1	SKFALIRKIGENI

## 25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 (SEQ ID NO: 594) shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) (SEQ ID NO: 600) from *N.gonorrhoeae*:

	orf141.pep	DFGISPVYLVWAAAFKHLSPWAADSVDVA	30
30	orf141ng	WNPAEPAVYTAVEALAGSPTPLVAHLFGQTDGIPPVYLVWAAAFKHLSPWAAHPYDAA	126
	orf141.pep	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141ng	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRSVVLHIGCIGLIPVAHFFNPAAAAFAAAGL	186
35	orf141.pep	VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP	140
	orf141ng	VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTA	246

An ORF141ng nucleotide sequence (SEQ ID NO: 599) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 600):

40	1	MPSEAVSARP LCEYLLHLAI RPFLTLMLT YTPPDARPPA KTHEKPWLLL
	51	LMAFAWLWPG VFSHDLWNPA EPAVYTAVEA LAGSPTPLVA HLFQTDGFI

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

101 PPVYLWVAAA FKHLSPWAA HPYDAARFAG VFFAVIGLTS CGFAGFNFLG  
151 RHHGRSVVLI HIGCIGLIPV AHFFNPAAAA FAAAGLVLHG YSLARRRVIA  
201 ASFLLGTGWT LMSLAAAYPA AFALMLPLPV LMFFRPWQSR RLMLTAVASL  
251 AFALPLMTVY PLLLAKTQPA LFAQWLNHYV FGTGGVRHI QRAFSLFHYL  
301 KNLLWFAPPG LPLAVWTVCR TRLFSTDWGI LGIVWMLAVL VLLAFNPQRF  
351 QDNLVWLLPP LALFGAAQLD SLRRGAAAFV NWFGIMAFGL FAVFLWTGFF  
401 AMNYGWPAKL AERAAFYSPY YVPDIDPIM AVAVLFTPLW LWAITRKNIR  
451 GRQAVTNWAA GVTLTWALLM TLFLPWLDAA KSHAPVVRSM EASFSPCLKR  
501 ELSDGIECIG IGGGDLHTRI VWTQYGTLPV RVGDVRCRYR IVRLPQNADA  
551 PQGWQTVWQG ARPRNKDSKF ALIRKIGENI LKTTD\*

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 601):

15  
20  
25  
30  
35  
40  
45

1 ATGCTGACCT ATACCCCGCC CGATGCCCCG CCGCCCCCCA AAACCCACGA  
51 AAAACCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGCTG TGGCCCGCGC  
101 TGTTTTCCCA CGATTTGTGG AATCCTGCCG AACCTGCCGT CTATACCGCC  
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCC TTGGTTGCCC ATCTGTTCCG  
201 TCAAACCGAT TTCGGCATA CCCCCGTGTA TCTTTGGGTT GCCGCCGAT  
251 TCAAACATTT GCTGTCGCCG TGGGCAGCCG ACCCGTATGA TGCCGCACGC  
301 TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCTT GCGGCTTTGC  
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTT GTTTTAATCC  
401 ATATCGGCTG TATCGGCTG ATTCCGGTTG CCCATTTCCT CAATCCgccc  
451 gccgcccgcct tTGCCGCGC CGGACTGGTG CTGCacggct actcgctgGC  
501 ACGCCGGCGC GTGATtgccc cctctTtccT GCTCGGTACG GGTTGGACGT  
551 TGATGTCGCT GGCGGCAGCT TATCCGGCGG CGTTTGCGCT GATGCTGCCC  
601 CTGCCCCTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT  
651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC  
701 CGCTGCTCtt gGCAAAAACG CAGCCCGCGC TGTTTGCGCA ATGGCTCAAC  
751 TATCACGTTT TCGGTACGt cggcgGCGTG CGGCACaTTC AGAggGCatT  
801 Cagtttgttt cactatctgA AAaatctgct ttggttcgca ccgccccggc  
851 TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CACGCCTGTT TTCGACCGAC  
901 TGGGGGATTT TGGGCATTGT CTGGATGCTT GCCGTTTTGG TGCTGCTCGC  
951 CTTTAATCCG CAGCGTTTTT AAGACAACCT CGTCTGGCTG CTGCCGCCGC  
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG  
1051 GCTTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGGCTGT TTGCCGTGTT  
1101 CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG  
1151 CCGAACGCGC CGCCTACTTC AGCCCGTATT ACGTTCCCGA CATCGATCCC  
1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT  
1251 TACCCGGAAC AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCAG  
1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTTCT GCCGTGGCTG  
1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGTT  
1401 TTCCCCGGAA TTA AACCGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA  
1451 TAGCGGCGCG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGACA  
1501 TTGCCGCACC GCGTCGGCGA TGTCCGTTGC CGCTACCGTA TCGTCCGCCT  
1551 GCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGCTG  
1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTTG CACTGATACG GAAAATCGGG  
1651 GAAAATATAT TAAAAACAAC AGATTGA

This corresponds to the amino acid sequence (SEQ ID NO: 602; ORF141ng-1):

50  
55

1 MLTYTPPDAR PPAKTHEKPW LLLLMAFAWL WPGVFSHDLW NPAEPAVYTA  
51 VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAKHLLSP WAADPYDAAR  
101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLIHIGCIGL IPVAHFLNPA  
151 AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP  
201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLN  
251 YHVFGTFGGV RHQRAFSLF HYLKNLLWFA PPGLPLAVWT VCRTRLFSTD  
301 WGILGIVWML AVLVLAFNP QRFQDNLVWL LPPLALFGAA QLDLRRGAA  
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP  
401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL

451 DAAKSHAPVV RSMEASFSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG  
 551 ENILKTTD\*

- 5 ORF141ng-1 (SEQ ID NO: 602) and ORF141-1 (SEQ ID NO: 596) show 97.5% identity in 553 aa overlap:

```

orff141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP
|
|
|
orff141-1      MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP
|
|
|
10 orff141ng-1.pep LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN
|
|
|
orff141-1      LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN
|
|
|
15 orff141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
|
|
|
orff141-1      FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
|
|
|
orff141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
|
|
|
orff141-1      GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTAVASLAFALPLMTVYPLLLAKT
|
|
|
20 orff141ng-1.pep QPALFAQWLNHYHVFQTFGGVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCTRRLFSTD
|
|
|
orff141-1      QPALFAQWLDYHVFQTFGGVRHVQTAFSLFYLYLKNLLWFALPALPLAVWTVCTRRLFSTD
|
|
|
orff141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
|
|
|
orff141-1      WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
|
|
|
25 orff141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAI TRK
|
|
|
orff141-1      FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAI TRK
|
|
|
orff141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMESFSPELKRELS DGIE
|
|
|
30 orff141-1      NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSMESLSPELKRELS DGIE
|
|
|
orff141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
|
|
|
orff141-1      CIGIGGGDLHTRIVWTQYGTLP HRVGDVQC RYRIVLLPQNADAPQGWQTVWQGARPRNKD
|
|
|
orff141ng-1.pep SKFALIRKIGENILKTTDX
|
|
|
35 orff141-1      SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 72

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 603):

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGGCA CAGCAATTGG
51  GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
5   151  AGCGGTTTTC AGGTAGGCTA TACGTTTTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 604; ORF142):

```

1  ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
10  51  SGFQVG YTF*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 605):

```

1  ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC
51  TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTC TATGTAAATT
15  101  ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAGTTT TGACGGCCAT
151  CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCTTTT
201  CGGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
251  CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT
301  ACTGATTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
351  CTATCTCGGT GTAAAACGTG GGATGAGGGA AACAAAAAGT TACATTGATG
20  401  ATGCCGAAC TACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA
451  CTTTCCCACA AAGAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA
501  ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG
551  CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
601  GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC
25  651  CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG
701  CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG
751  TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA
801  ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC
851  AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGTCGGCAC AGCAATTGGG
30  901  ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC
951  CCGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAG
1001 GCGGTTTTCA GGTAGGCTAT ACGTTTTTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 606; ORF142-1):

```

35  1  MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH
51  RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQA VSGLSE VYDYNKSYN
101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
151 LSHKEYIGRS TADFKLKYKR GTGMKDALRA PEEAFGEGTS RMKIWTASAD
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
40  251  SAERGWIWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAG
301  IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF142 (SEQ ID NO: 604) shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) (SEQ ID NO: 608) from *N.gonorrhoeae*:



-435-

```

    orf142.pep                               QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY   30
    orf142ng      RGWYWRNDLSWQFKPGHQLYLGAADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313

5    orf142.pep      DIFTGRALKKPEFFQSRKWASGFQVGYTF      59
    orf142ng      DIFTGRALKKPEYFQTKKWVTGFQVGYSF      342

```

The complete length ORF142ng nucleotide sequence (SEQ ID NO: 607) is:

```

10      1  ATGGATAATT  CGGGTAGTGA  GGCGACAGGA  AAATACCAAG  GAAATATCAC
      51  TTTCTCTGCC  GACAATCCTT  TTGGACTGAG  TGATATGTTC  TATGTAAATT
     101  ATGGACGTTC  AATTGGCGGT  ACGCCCGATG  AGGAAAATTT  TGACGGCCAT
     151  CGCAAAGAAG  GCGGATCAAA  CAATTACGCC  GTACATTATT  CAGCCCCTTT
     201  CGGTAATGG  ACATGGGCAT  TCAATCACAA  TGGCTACCGT  TACCATCAGG
     251  CGGTTTCCGG  ATTATCGGAA  GTCTATGACT  ATAATGGAAA  AAGTTACAAC
    15      301  ACTGATTTCG  GCTTCAACCG  CCTGTGTAT  CGTGATGCCA  AACGCAAAAC
     351  CTATCTCAGT  GTAAACTGT  GGACGAGGGA  AACAAAAAGT  TACATTGATG
     401  ATGCCGAAct  GACTGTACAA  CGGCGTAAAA  CCACAGGTTG  GTTGGCAGAA
     451  CTTTCCCA  AAGGATATAT  CGGTCGCAGT  ACGGCAGATT  TTAAGTTGAA
     501  ATATAAACAC  GGCACCGGCA  TGAAAGATGC  TCTGCGCGCG  CCTGAAGAAG
    20      551  CCTTTGGCGA  AGGCACGTCA  CGTATGAAAA  TTTGGACGGC  ATCGGCTGAT
     601  GTAAATACTC  CTTTTCAAA  CGGTAAACAG  CTATTTGCCT  ATGACACATC
     651  CGTTCATGCA  CAATGGAACA  AAACCCCGCT  AACATCGCAA  GACAAACTGG
     701  CTATCGGCGG  ACACCAcACC  GTACGTGGCT  TCGACGGTGA  AATGAGTTTG
     751  CCTGCCGAGC  GGGGATGGTA  TTGGCGCAAC  GATTTGAGCT  GGCAATTTAA
    25      801  ACCAGGCCAT  CAGCTTTATC  TTGGGGCTGA  TGTAGGACAT  GTTTCAGGAC
     851  AATCCGCCAA  ATGGTTATCG  GGCCAAACTC  TAGCCGGCAC  AGCAATTGGG
     901  ATACGCGGGC  AGATAAAGCT  TGGCGGCAAC  CTGCATTACG  ATATATTTAC
     951  CGGCCGTGCA  TTGAAAAAGC  CCGAATATTT  TCAGACGAAG  AAATGGGTAA
    30      1001  CGGGGTTTCA  GGTGGGTTAT  TCGTTTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 608):

```

35      1  MDNSGSEATG  KYQGNITFSA  DNPFGLSDMF  YVNYGRSIGG  TPDEENFDGH
     51  RKEGGSNNYA  VHYSAPFGKW  TWAFNHNGYR  YHQA VSGLSE  VYDYNKSYN
     101  TDFGFNRLLY  RDAKRKTYLS  VKLWRETAKS  YIDDAELTVQ  RRKTGWLAE
     151  LSHKGYIGRS  TADFKLKYKH  GTGMKDALRA  PEEAFGEGTS  RMKIWTASAD
     201  VNTPFQIGKQ  LFAYDTSVHA  QWNKTPLTSQ  DKLAIGGHHT  VRGFDGEMSL
     251  PAERGWYWRN  DLSWQFKPGH  QLYLGADVGH  VSGQSAKWLS  GQTLAGTAIG
     301  IRGQIKLGGN  LHYDIFTGRA  LKKPEYFQTK  KWTGFGQVGY SF*

```

40 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng (SEQ ID NO: 608) and ORF142-1 (SEQ ID NO: 606) show 95.6% identity over 342aa overlap:

```

45    orf142-1.pep  MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
    orf142ng-1    MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA

    orf142-1.pep  VHYSAPFGKWTWAFNHNGYRYHQA VSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLG

```

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558  
Score = 119 bits (295), Expect = 3e-26  
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)

Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYGKSYNTDFGFNRLLYRDAKRKTYLSV 121  
+S P+G W +N++ RY + G S F +R+++RD KT ++  
Sbjct: 281 -FSMPYGYWNLGYNYSQSRYRNTFINRDFPWHSTGDSDFHRSLSRVVFRDGMTKTAIAG 339

Query: 182 EEA FGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHTTV 241  
 +++ E + WT SA P Y S++ Q++ L ++L +GG ++  
 Sbjct: 400 DKSADEPRAEFNKWTLASYYHPV---TDSITYLGSLYGQYSARALYGSEQLTLGGESSI 456

35 Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWVTGFQVGYSF 342  
 A+G+ + L + G + P + Q V G++VG SF  
 Subject: 516 GAVGMTVASRW---LSQQVTVGWPIISYPAWLQPDMTMVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 73**

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 609):

```

1  ATGCGGACGA AATGGTCAGC AGTGAGAAGC TGCTTACTTG GgCGGACACC
51  GCCGACATCG ATACCGCTTT GAACCTGTTG TACCGTTTGC AAAAACTCGA
101 ATTCTCTAT  GCGCATGAAA ACGGTCATTG AGACGGCATC AATTTGwCGG
151 ACGAGCAATT GCCGTTGCTG ATGGAACAAT TGTCCGGCAG CCGTAAGGCG
201 TTATTGGTCG ATCGGAACGG TCTGTATCTT GCCAACGCCA ATTTCCATCA
251 TGAGGCGGCG GAAGAGTTGG GGTGTGTGGC GGCAGAAGTC GCACAGATGG
301 AAAAGAAATA CCGGCTGCTG ATTAAGAACA AC..

```

This corresponds to the amino acid sequence (SEQ ID NO: 610; ORF143):

```

1  MRTKWSÄVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
51  EQLPLLMEQL SGSGKALLVD RNGLYLANAN FHHEAAEELG LLAAEVAQME
101 KKYRLLIKNN ..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 611):

```

1  ATGGAATCAA CACTTTCACT ACAAGCAAAT TTATATCCCC GCCTGACTCC
51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA
101 CTTTGTGTGA CAGCCTGTTG AAAGCAGATG CGGACGAAAT GGTGAGCAGT
151 GAGAAGCTGC TTACTTGGGC GGACACCGCC GACATCGATA CCGCTTTGAA
201 CCTGTTGTAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG
251 GTCATTGAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
301 GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
351 GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
401 TGTTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCTGATT
451 AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
501 CGGTCAGAGC GAATTGACAT TTTTCCCATT GTATATCGGT TCAACCAAAT
551 TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT
601 ACTTTGGTAA GGATTTTATA CCGCCGTTAC AGCAACCGCG TGTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 612; ORF143-1):

```

1  MESTLSLQAN LYPRLTPAGA FYAVSSDAPS AGKTLHSLL KADADEMVSS
51  EKLLTWADTA DIDTALNLLY RLQKLEFLYG DENGHSDGIN LSDEQLPLLM
101 EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLLI
151 KNNLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
201 TLVRILYRRY SNRV*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF143 (SEQ ID NO: 610) shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) (SEQ ID NO: 614) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF143a nucleotide sequence (SEQ ID NO: 613) is:

	1	ATGGAATCAA	CANTTTCACT	ACAAGCAAAT	TTATATCNCC	GCCTGACTCC
20	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGNCCCCAGT	GCCGGTAAAA
	101	CTTTGTTGCA	CAGCCTGTTG	AAAGCGGATG	CGGACGAAAT	GGTNAGCAG
	151	GAGAAGCTGC	TTACCTGGGC	GGANACCGCC	GACATCGATA	CCGCTTTGAA
	201	CCTGTTGTAC	CGTTTGCAAA	AACTCGAATT	CCTCATGGC	GATGAAAAAG
	251	GTCAATCAGA	CGGCATCAAT	TTGTCGGACG	AGCAATTGCC	GTTGCTGATG
25	301	GAACAATTGT	CCGGCAGCGG	TAAGGCGTTA	TTGTCGATC	GGAACGGTCT
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GGCGGCGGAA	GAGTTGGGGT
	401	TGTTGGCGGC	AGAAGTCGCA	CAGATGGAAG	AGAAATACCG	GCTGCNNATT
	451	AAGAACAACC	TGTATATCAA	CAATAACGCT	TGGGGCGTTT	GCGATCCTTC
	501	CGGTCAGAGC	GAATTGACAT	TTTTCCCAT	GTATATCGGT	TCAACCAAAT
	551	TTATTTTGGT	TATCGGCGGC	ATTCCTCGAT	TGGGCAAGA	GGCATTGTGT
30	601	ACTTTGGTAA	GGATNTTATA	CCNCCNGTTA	CAGCAACCGC	GTGTAAAACT
	651	TGGGAGAGAG	GANGGGTTAT	GCAGCAATTA	TTGA	

This encodes a protein having amino acid sequence (SEQ ID NO: 614):

35

1	MESTXSLQAN	LYXRLTPAGA	FYAVSSDXPS	AGKTLHLSLL	KADADEMVSS
51	EKLLTWAXTA	DIDTALNLLY	RLQKLEFLYG	DENGHSDGIN	LSDEQLPLLM
101	EQLSGSGKAL	LVDRNGLYLA	NANFHHEAAE	ELGLLAAEVA	QMEKKYRLXI
151	KNNLYINNNA	WGVCDPSEQS	<u>ELTFFPLYIG</u>	<u>STKFILVIG</u>	IPDLGKEAFV
201	TLVRXLYXXL	QOPRVKLGRE	XGLCSNY *		

40 ORF143a (SEQ ID NO: 614) and ORF143-1 (SEQ ID NO: 612) show 97.1% identity in 207 aa overlap:

[illegible]

```

orf143-1      NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPSGQSELTFFPLYIG
orf143a.pep   STKFILVIGGIPDLGKEAFVTLVRXLY
              |||||
orf143-1      STKFILVIGGIPDLGKEAFVTLVRILY

```

## 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 (SEQ ID NO: 610) shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) (SEQ ID NO: 616) from *N.gonorrhoeae*:

```

10 orf143.pep   MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLXDEQLPLLMEQL   60
      |||||
orf143ng      MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQL   60

orf143.pep   SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIKNN   110
      |||||
orf143ng      SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIIRNNLYINNNAWGV   120

```

15 An ORF143ng nucleotide sequence (SEQ ID NO: 615) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 616):

```

20 1  MRTKWSAVRS  CSRADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDGINLSD
    51  EQLPLLMEQL  SGSGKALLVD  RNGLYLANAN  FHESAEELG  LLAAEVAQME
   101  KKYRLLIRNN  LYINNNAWGV  CDPSGQSELT  FFPLYIGSTK  FILVIAGIPD
   151  LSKGGICYFG  KDFIPPLQQP  RVKLGTTGIM  RQLLISILED  LNNTSTDIIA
   201  SAVISTDGLP  MATMLPSHLN  SDRVGAISAT  LLALGSRVQ  ELACGELEQV
   251  MIKKGSGYIL  LSQAGKDAVL  VLVAKETGRL  GLILLDAKRA  ARHIAEAI*

```

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 617):

```

25 1  ATGGAATCAA  CACTTTCAC  ACAAGCGAAT  TTATATCCCT  GCCTGACTCC
    51  TGCCCGTGCA  TTTTATGCCG  TATCCAGCGA  TGCCCCAGT  GCCGGTAAAA
   101  CTTTGTGCG  CAGCCTGTTG  AAAGCGGATG  CGGACGAAGT  GGTCAGCAGT
   151  GAGAAGCTGC  TCGCGCGGGA  CACCGCCGAC  ATCGATACCG  CTTGAAACCT
   201  GTTGTACCGT  TTGCAAAAAC  TCGAATTCCT  CTATGGCGAT  GAAAACGGTC
   251  ATTCAGACGG  CATCAATTG  TCGGACGAGC  AATTGCCGTT  GCTGATGGAA
   301  CAATTGTCCG  GCAGCGGTAA  GGCATTATTG  GTCGATCGGA  ACGGTCTGTA
   351  TCTTGCCAAC  GCCAATTTCC  ATCATGAGTC  GGCGGAAGAG  TTGGGGTTGT
   401  TGGCGGCAGA  AGTCGCACAG  ATGGAAAAGA  AATACCGGCT  GCTGATTAGG
   451  AACAACTGT  ATATCAACAA  TAACGCTTGG  GGCGTTTGGC  ATCCTTCCGG
   501  TCAGAGCGAA  TTGACATTTT  TCCCATTTGA  TATCGGTTCA  ACCAAATTTA
   35 551  TTTTGGTTAT  CGCCGCGATT  CCCGATTGTA  GCAAAGAGGC  ATTTGTTACT
    601  TTGTAAGGA  TTTTATACCG  CCGTTACAGC  AACCGCGTGT  AA

```

This corresponds to the amino acid sequence (SEQ ID NO: 618; ORF143ng-1):

```

40 1  MESTLSIQAN  LYPCLTPAGA  FYAVSSDAPS  AGKTLRLSLL  KADADEVVSS
    51  EKLLAADTAD  IDTALNLLYR  LQKLEFLYGD  ENGHSDGINL  SDEQLPLLME
   101  QLSGSGKALL  VDRNGLYLAN  ANFHESAEEL  LGLLAAEVAQ  MEKKYRLIR
   151  NNLYINNNAW  GVCDSGQSE  LTFFFLYIGS  TKFILVIAGI  PDLSKEAFVT
   201  LVRILYRRYS  NRV*

```

ORF143ng-1 (SEQ ID NO: 618) and ORF143-1 (SEQ ID NO: 612) show 95.8% identity in 214 aa overlap:

```

5      orf143ng-1.pep MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLLRSLKADADEVVSSEKLLA-ADTA  59
      orf143-1      MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLLHSLKADADEMVSSEKLLTWADTA  60
      orf143ng-1.pep DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 119
      orf143-1      DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 120
10     orf143ng-1.pep NANFHHESAEEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFFPLYIG 179
      orf143-1      NANFHHESAEEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPSGQSELTFFFPLYIG 180
      orf143ng-1.pep STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV  213
      orf143-1      STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV  214
15

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 74

20 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 619):

```

      1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGr
101  CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
25  151  ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCG ACCGTGGTTC
      201  GGATTCTGTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
      251  ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
      301  ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
      351  GACGATAGAC AATACGTTCA ACCGCATCTG GaCGGGTCAA wTyCCAGCGT
30  401  CCGTGGATG. .

```

This corresponds to the amino acid sequence (SEQ ID NO: 620; ORF144):

```

      1  MTFLQRLQGL ADNKICAFaw FVRRFDEER VPQXAASMTF TTLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP XGADMVFDYI NAFREQANRL
35  101  TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM. . .

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 621):

```

      1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
40  101  CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
      151  ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCG ACCGTGGTTC
      201  GGATTCTGTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG

```

-441-

5  
10  
15  
20

```

251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGACG CTTTGTCTGT GGGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCCTCG GCGGCAGGCG TTTGTCGGGG CTTTGGCAAC
651 AGCGTTTTGT CTGAAACCG CGCGCTCCCT CTTCACTTGG TATATGGGCA
701 ATTTGACGG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAGAAGG CAAAGCCTTG CCTGTTTCTG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 GCGGCGACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACGCGAGTA G

```

This corresponds to the amino acid sequence (SEQ ID NO: 622; ORF144-1):

25  
30

```

1  MTFLLQRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIIV QGADMVFDYI NAFREQANRL
101 TAIGSVMLLV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFTMT LLLWGLYRFV
201 PNRFPVARQA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLAHGYIY SGRQGVVLT
351 GADSIELNEL FKLFFVRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF144 (SEQ ID NO: 620) shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) (SEQ ID NO: 624) from strain A of *N. meningitidis*:

40  
45

```

          10      20      30      40      50      60
orf144.pep MTFLLQRLQGLADNKNKICAFWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
          |||||
orf144a    MTFLLQRLQGLADNKNKICAFWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
          10      20      30      40      50      60

          70      80      90     100     110     120
orf144.pep PVFDRWSDSFVSFVNQTIIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
          |||||
orf144a    PVFDRWSDSFVSFVNQTIIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
          70      80      90     100     110     120

          130
orf144.pep NTFNRIWRVXXQRPWM
          |||||

```

orf144a      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

130                  140                  150                  160                  170                  180

The complete length ORF144a nucleotide sequence (SEQ ID NO: 623) is:

5	1	ATGACCTTTT	TACAACGTTT	GCAAGGTTTG	GCAGACAATA	AAATCTGTGC
	51	GTTTGCATGG	TTCGTCGTCC	GCCGCTTTGA	TGAAGAACGC	GTACCGCAGG
	101	CGGCGGCAAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTGCTG
	151	ACCGTGATGG	TGGCGGTGCG	TTCGATTTTC	CCCGTGTTTCG	ACCGNTGGTC
	201	GGATTCTGTT	GTCTCCTTCG	TCAACCAAA	CATTGTGCGC	CAGGGCGCGG
10	251	ACATGGTNTT	CGACTATATC	AATGCGTTC	GCGAGCAGGC	GAACGCGCTG
	301	ACGGCAATCG	GCAAGCGTAT	GCTGGTCTGT	ACCTCGCNGA	TGCTGATTCTG
	351	GACGATAGAC	AATACGTTCA	ACCGCATCTG	GCGGGTCAAT	TCCCAGCGTC
	401	CGTGGATGAT	GCAGTTTCTC	GTCTATTGGG	CTTTACTGAC	GTTCGGGCCG
	451	CTGTCTTTGG	GCGTGGGCAT	TTCTTTTATN	GTCCGCTCGG	TACAGGATGC
15	501	CGCGCTTGCC	TCAGGTGCGC	CGCAGTGGTG	GGGCGCGTTG	CGAACGGCGG
	551	CGACGCTGAN	CTTCATGACG	CTTTTGCTGT	GGGGCTGTGA	CCGCTNCGTG
	601	CCAAACCGCT	TCGTTCCCGC	GCGGCANGCG	TTTGTCTGGG	CTTTGCGAAC
	651	AGCGTTCTGT	CTGGAAACCG	CGCGTTCCTT	CTTTACTTGG	TATATGGGCA
	701	ATTTTCGACG	CTACCGCTCG	ATTTACGGNG	CGTTTGCCGC	CGTGCCGTTT
20	751	TTTCTGTTGT	GGCTGAACTT	GTTGTGGACG	CTGGTCTTGG	GCGGCGCGGT
	801	GCTGACTTCT	TCACTCTCCT	ACTGGCAGGG	AGAAGCGTTC	CGCAGGGNCT
	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAACTCT	GCTGCTTCTG
	901	GATCGGCGCG	AAAAAGAAGG	CNAAGCCTTG	CCTGTTTCAAG	AGTTCAGACG
	951	GCATATCAAT	ATGGGCTACG	ACGAGTTGGG	CGAGCTTTTG	GAAAAGCTGG
25	1001	CGCGGCACGG	CTACATCTAT	TCCGGCAGAC	AGGGTTGGGT	GTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAACGAACTC	TTCAAGCTCT	TCGTTTACCG
	1101	TCCGTTGCCT	GTGGAAAGGG	ATCATGTGAA	CCAAGCTGTC	GATGCGGTAA
	1151	TGATGCCGTG	TTTGACAGCT	TTGAACATGA	CGCTGGCAGA	GTTTGACGCT
	1201	CAGGCGAAAA	AACAGCAGCA	ATCTTGA		
30						

This encodes a protein having amino acid sequence (SEQ ID NO: 624):

	1	<u>MTFLQRLQGL</u>	<u>ADNKICAFAW</u>	<u>FVVRFRDEER</u>	<u>VPQAAASMTF</u>	<u>TTLALVPVL</u>
	51	<u>TVMVAVASIF</u>	<u>PVFDRWSDSF</u>	<u>VSFVNQTIVP</u>	<u>QGADMVFDYI</u>	<u>NAFREQANRL</u>
35	101	<u>TAIGSVMLVV</u>	<u>TSXMLIRTID</u>	<u>NTFNRIWRVN</u>	<u>SQRPWMQFL</u>	<u>VYWALLTFGP</u>
	151	<u>LSLGVGISFX</u>	<u>VGSVQDAALA</u>	<u>SGAPQWSGAL</u>	<u>RTAATLXFMT</u>	<u>LLWGLYRXV</u>
	201	<u>PNRFVPARKA</u>	<u>FVGALATAFC</u>	<u>LETARSLFTW</u>	<u>YMGNFDDGYS</u>	<u>IYGAFAAVPF</u>
	251	<u>FLLWLNLLWT</u>	<u>LVLGGAVLTS</u>	<u>SLSYWQGEAF</u>	<u>RRXFDSRGRF</u>	<u>DDVLKILLLL</u>
	301	<u>DAAQKEGXAL</u>	<u>PVQEFRRHIN</u>	<u>MGYDELGELL</u>	<u>EKLARHGYIY</u>	<u>SGRQGWVLKT</u>
	351	<u>GADSIELNEL</u>	<u>FKLFVYRPLP</u>	<u>VERDHVNQAV</u>	<u>DAVMMPCLQT</u>	<u>LNMTLAEFDA</u>
40	401	OAKKQOQS*				

ORF144a (SEQ ID NO: 624) and ORF144-1 (SEQ ID NO: 622) show 97.8% identity in 406 aa overlap:

45	orf144a.pep	MTFLQLRQLGLADNKICAFAWFVRRFDEERVQAAASMTFTLLALVPVLTVMVAVASIF
	orf144-1	MTFLQLRQLGLADNKICAFAWFVRRFDEERVQAAASMTFTLLALVPVLTVMVAVASIF
	orf144a.pep	PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVT SXMLIRTID
	orf144-1	PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
50	orf144a.pep	NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL



	orf144-1	NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL	
	orf144a.pep	RTAATLXFMTLLWGLYRXVPNRFVPARXAFVFGALATAFCLETARSLFTWYMGNFDDGYRS	
	orf144-1	RTAATLTFMTLLWGLYRFVFNRFVPARQAFVFGALATAFCLETARSLFTWYMGNFDDGYRS	
5	orf144a.pep	IYGAFAAVPPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL	
	orf144-1	IYGAFAAVPPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL	
	orf144a.pep	DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIENEL	
10	orf144-1	DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLKTGADSIENEL	
	orf144a.pep	FKLFVYRPLPVERDHVNQAVDAVMMPCLOTLNMTLAEFDAQAKKQQQS	408
	orf144-1	FKLFVYRPLPVERDHVNQAVDAVMTPCLOTLNMTLAEFDAQAKKRQ	406

### Homology with a predicted ORF from *N.gonorrhoeae*

- 15 ORF144 (SEQ ID NO: 620) shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) (SEQ ID NO: 626) from *N.gonorrhoeae*:

	orf144.pep	MTFLQRLQGLADNKICAFWFVRRFDEERVQPXAAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWFVIRRFSEERVQAAAASMTFTTLLALVPVLTVMVAVASIF	60
20	orf144.pep	PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
25	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence (SEQ ID NO: 625) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 626):

30	1	MTFLQCWQGS	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLVV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	CLETARFLFTW	YMGNFDDGYRS	IYGAFAAVPP
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
35	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTPCLOT	LNMTLAEFDA
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 627):

40	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTGCATGG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG

5	151	ACCGTAATGG	TCGCGGTGCG	TTCGATTTTC	CCCGTGTTCC	ACCGCTGGTC
	201	GGATTCGTTC	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGGCGCGG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
	301	ACCGCCATCG	CGACCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCC
	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	CGGGGTTAAC	ACGCAACGCC
10	401	CCTGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCCT
	451	TTGTCTTTGG	GTGTGGGCAT	TTCTTTTATG	GTCCGGTCGG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACCGGTTG	AAGACGGCGG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGCGC	TTTGTCCGAG	CTTTGATTAC
15	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCCT	GTTACCTGG	TATATGGGCA
	701	ATTTGACGCG	CTACCGCTCG	ATTTACGGCG	CATTGCGCG	CGTGCCGTTT
	751	TTCTGTCTGT	GGTTAAACCT	GCTGTGGACG	CTGTCTTGG	GCGGGGCGGT
	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTCT	CGCAGGGGAT
	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
20	901	GATGCGCGCG	AAAAAGAAAG	CCGAACCTGT	TCCGTTACAG	AGTTTCAGACG
	951	GCATATCAAT	ATGGGTTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
20	1151	TGAgcgctg	TTTGCAGCT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

25 This encodes a variant of ORF144ng, having the amino acid sequence (SEQ ID NO: 628;  
ORF144ng-1):

30

1	<u>MTFLQRWQGL</u>	ADNKICAFAW	FVIRRFSEER	VPQAAASMTF	TTLALVPVL
51	<u>TVMVASIF</u>	PVFDRWDSF	VSVNQITVP	QGDMVFDYI	DAFRDQANRL
101	<u>TAIGSVMLVV</u>	TSLMLIRTID	NAFNRIWRVN	TQRPWMQFL	VYWALLTFGP
151	<u>LSLGVGSFM</u>	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
201	<u>PNRFVPARQA</u>	FVGALITAF	LETARFLFTW	YMGNFDDGYS	IYGAFAAVPF
251	<u>FLLWLNLLWT</u>	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
301	<u>DAAQKEGRTL</u>	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRGQWVLTK
351	<u>GADSIELSEL</u>	FKLFVYRPLP	VERDHVNQAV	DAVMTPCLQT	LNMTLAEFDA
401	<u>QAKKQQQS*</u>				

ORF144ng-1 (SEQ ID NO: 628) and ORF144-1 (SEQ ID NO: 622) show 94.1% identity in 406 aa overlap:

	orf144ng-1.pep	MTFLQRWQGLADNKICAFAWFVIRRFSEERVVPQAAASMTFTTLLALVPVLTVMVAVASIF
	orf144-1	MTFLQRLQGLADNKICAFAWFVVRRFDEERVVPQAAASMTFTTLLALVPVLTVMVAVASIF
40	orf144ng-1.pep	PVFDWRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
	orf144-1	PVFDWRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
45	orf144ng-1.pep	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDVLSSGAQQWADAL
	orf144-1	NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
	orf144ng-1.pep	KTAARLAFMTLLWGLYRFVPNRFVPARQAFVGALITAFCLETARFLFTWYMGNFDDGYRS
	orf144-1	RTAATLTFMTLLWGLYRFVPNRFVPARQAFVGALATAFCLETARSLFTWYMGNFDDGYRS
50	orf144ng-1.pep	IYGAFAAVPPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILL

```

      |||
orfl44-1  IYGAF AAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
      |||

orfl44ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL
      |||:::|
5 orfl44-1  DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYSGRQGWVLKTGADSIELNEL
      |||:::|

orfl44ng-1.pep FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKQQQS
      |||
orfl44-1  FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKRQ
      |||

```

- 10 On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 75

- 15 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 629):

```

      1  ..AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
      51  AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
      101 GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
      151 ACCCGCCGCA AATGGCTGGA TGCCACAGAA CGCCAACACC TGCGCCAAG
20 201  CCTGCTTGAA ACACGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 630; ORF146):

```

      1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
      51  TRRKWLDAHE RQHLRQSLLE TREHG*
25

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 631):

```

      1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
      51  CGAACGCTAC CGTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG
      101 CCGTCCTGTT CGCCACCGCC TCCGCCGGC TGCTCCACCT CCAACACGGC
      151 GAGTGGATAG GGATACCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTTCA
      201 AGGGGCGATT TACTCAAGG CGGTGGAAACG TATGCTCGGC ACGGTCATCG
      251 GGCTGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
      301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
      351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
      401 CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
      451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
      501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
      551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
      601 AGGCGCATGA CCCGCGAAGC CCTCGAGGAG AACATGGCGA AAATGCGCCA
      651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCATCTCGCC GCCACATCGG
      701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCCAC
      751 CGTAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
      801 GCAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
      851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
45 901  AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA

```

951	AGCCCTCGCC	GAACACCTCC	ACTACCAATG	GCAGGGCTTC	CTCTGGCTCA
1001	GCACCAATAT	GCGTCAGGAA	ATTTCCGCC	TCGTATCCT	GCTGCAACGC
1051	ACCCGCCGCA	AATGGCTGGA	TGCCCACGAA	CGCCAACACC	TGCGCCAAAG
1101	CCTGCTTGAA	ACACGGGAAC	ACGGCTGA		

This corresponds to the amino acid sequence (SEQ ID NO: 632; ORF146-1):

1	MNTSQRNRLV	SRWLNSYERY	RYRRLIHAVR	LGGAVLFATA	SARLLHLQHG
51	EWIGMTVFVV	LGMLQFQAI	YSKAVERMLG	TVIGLGAGLG	VLWLNQHYFH
101	GNLLFYLTVG	TASALAGWAA	VGKNGYVPM	AGLTMCM	DNGSEWLD
151	LMRAMNVLIG	AAIATAAAKL	LPLKSTLMWR	FMLADNLADC	SKMTAEISNG
201	RRMTREERLEE	NMAKMRQINA	RMVKSRS	ATSGESRISP	AMMEAMQHAH
251	RKIVNTTELL	LTAAKLQSP	KLNGSEIRLL	DRHFTLLQTD	LQQTVALING
301	RHARRIRIDT	AINPELEALA	EHLHYQWQGF	LWLSTNMRQE	ISALVILLQR
351	TRRKWLDAHE	ROHLROSILLE	TRHG*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF146 (SEQ ID NO: 630) shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) (SEQ ID NO: 634) from strain A of *N. meningitidis*:

```

                                10      20      30
orf146.pep                      RHARRIRIDTAINPELEALAEHLHYQWQGF
                                |||
orf146a      KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF
                280      290      300      310      320      330

                                40      50      60      70
orf146.pep      LWLSTDMRQEISALVILLQTRRKWLDAHERQHLRQSLLETREHGX
                |||||:|||||
orf146a      LWLSTNMRQEISALVILLQTRRKWLDAHERQHLRQSLLETREHSX
                340      350      360      370

```

The complete length ORF146a nucleotide sequence (SEQ ID NO: 633) is:

1	ATGAACACCT	CGCAACGCAA	CCGCTCTGTC	AGCCGCTGGC	TCAACTCCTA
51	CGAACGCTAC	CGCTACCGCC	GCCTCATCCA	CGCCGTCCGG	CTCGGCGGGG
101	CCGTCTGTGT	CGCCACCGCC	TCCGCCCGGC	TGCTCCACCT	CCAACACGGC
151	GAGTGGATAG	GGATGACCGT	CTTCGTCGTC	CTCGGCATGC	TCCAGTTTCA
201	AGGGGCGGATT	TACTCCAAGG	CGGTGGAACG	TATGCTCGGG	ACGGTCATCG
251	GGCTGGCGCG	GGGTTTGGGC	GTTTTATGGC	TGAACACAGCA	TTATTTCCAC
301	GGCAACCTCC	TCTTCTACCT	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG
351	CTGGGCGGCG	GTCGGCAAAA	ACGGCTACGT	CCCTATGCTG	GCGGGGCTGA
401	CGATGTGCAT	GCTCATCGGC	GACAACGGCA	GCGAATGGTT	CGACAGCGGC
451	CTGATGCGCG	CGATGAACGT	CCTCATCGGC	CGGGCCATCG	CCATCGCGGC
501	CGCCAAACTG	CTGGCGCTGA	AATCCACACT	GATGTGGCGT	TTCATGCTTG
551	CCGACAACTT	GACCGATGTC	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC
601	AGGCGCATGA	CCCGCGAACG	CCTCGAAGAG	AACATGGCGA	AAATGCGCCA
651	AATCAACGCA	CGCATGTGTA	AAAGCCGCAG	CCACTCTGCC	GCACATCGG
701	GCGAAAGCCG	CATCAGCCCC	GCCATGTATG	AAGCCATGCA	GCACGCCCC
751	CGTAAATTTG	TCAACACCAC	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT
801	GCAATCTCCC	AAACTCAACG	CGAGCGAAAT	CCGGTCTGCTT	GACGCCCACT

5  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

This encodes a protein having amino acid sequence (SEQ ID NO: 634):

10  
 1 MNTSQRNRLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG  
 51 EWIGMTVFVV LGMLQFQGA IYKKAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWFD SG  
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG  
 201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH  
 251 RKIVNTTELL LTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING  
 15 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR  
 351 TRRKWLDAHE RQHLRQSLLE TREHS\*

ORF146a (SEQ ID NO: 634) and ORF146-1 (SEQ ID NO: 632) show 99.5% identity in 374 aa overlap:

20 orf146a.pep MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV  
 orf146-1 MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV  
 orf146a.pep LGMLQFQGA IYKKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA  
 25 orf146-1 LGMLQFQGA IYKKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA  
 orf146a.pep VGKNGYVPM LAGLTMCLIGDNGSEWFD SGLMRAMNVLIGAAIAIAAAKL LPLKSTLMWR  
 orf146-1 VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKL LPLKSTLMWR  
 30 orf146a.pep FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP  
 orf146-1 FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP  
 orf146a.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING  
 orf146-1 AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING  
 35 orf146a.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE  
 orf146-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTRRKWLDAHE  
 orf146a.pep RQHLRQSLLE TREHSX  
 40 orf146-1 RQHLRQSLLE TREHGX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF146 (SEQ ID NO: 630) shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) (SEQ ID NO: 636) from *N.gonorrhoeae*:

```

orf146.pep                                RHARRIRIDTAINPELEALAEHLHYQWQGF 30
|||||
orf146ng      KLNGSEIRLLDRHFTLLQTDLQQTAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364

orf146.pep      LWLSTDMRQEISALVILLQTRRRKWLDAHERQHRLRQSLLETREHG 75
|||||:|||||
orf146ng      LWLSTNMRQEISALVIPLQTRRRKWLDAHERQHRLRQSLLETREHG 409

```

An ORF146ng nucleotide sequence (SEQ ID NO: 635) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 636):

```

10      1  MSGVRFPSPA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
      51  YERYRHRLI  HAVRLGGTVL FATALARLLH LQHGEWIGMT VFVVLGMLQF
     101  QGAISNAVE  RMLGTVIGLG AGLGVWLNQ  HYFHGNLLFY LTIGTASALA
     151  GWAAVGKNGY VPMLAGLTMC MLIGDNGSEW LDSGLMRAMN VLIGAAIAIA
     201  AAKLLPLKST LMWRFMLADN LADCSKMAIE ISNGRRMTRE RLEQNMVKMR
15      251  QINARNVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLTTAAK
     301  LQSPKLNSE  IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
     351  EALAEHLHYQ WQGFVLWSTN MRQEISALVI PLQTRRRKWL DAHERQHRLRQ
     401  SLLETREHG*

```

20 Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 637):

```

      1  ATGAACCTCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
     51  CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCggaa
    101  cegtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAacacggc
    151  gAATGGATAG GGatgaCCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTCCA
25      201  AGGCcgatt tActccaacg cggtgGAacg taTGctcggc acggtcatcg
     251  ggctgGGCGC GGGTTTGgGc gTTTTATGGC TGAACCAGCA TTAtttcac
     301  ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccg
     351  ctGGGCGGCG GTCGGCAAAA acggetacgt ccctatgctg GCGGGGctgA
     401  CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
30      451  CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC
     501  CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
     551  CCGACAACCT GGCCGACTGC AGCAAAATGA TTCCCGAAAT CAGCAACGGC
     601  AGGCGTATGA CGCGCGAACG TTTGGAGCAG AATATGGTCA AAATGCGCCA
     651  AATCAACGCA CGCATGGTCA AAAGCCGCAG CCACCTCGCC GCCACATCGG
35      701  GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCCCAC
     751  CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
     801  GCAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGTGCTC GACCGCCACT
     851  TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
     901  AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
40      951  AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
    1001  GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
    1051  ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
    1101  CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

45 This corresponds to the amino acid sequence (SEQ ID NO: 638; ORF146ng-1):

```

      1  MNSSQRKRLS GRWLSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
     51  EWIGMTVFVV LGMLQFQGAI YSNAVERMLG TVIGLGAGLG VLWLNQHYFH
    101  GNLLFYLTIG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWLDSE
    151  LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
50      201  RRMTRERLEQ NMVKMRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH
     251  RKIVNTTELL LTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTAAALING
     301  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR

```

351 TRRKWLDAHE RQHLRQSLLE TREHG\*

ORF146ng-1 (SEQ ID NO: 638) and ORF146-1 (SEQ ID NO: 632) show 96.5% identity in 375 aa overlap

```

5      orf146-1.pep  MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
      orf146ng-1    MNSSQKRKLSGRWLNSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV

      orf146-1.pep  LGMLQFQGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVTGTASALAGWAA
10     orf146ng-1    LGMLQFQGAIYSNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA

      orf146-1.pep  VGKNGYVPMLAGLTMCMMLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
      orf146ng-1    VGKNGYVPMLAGLTMCMMLIGDNGSEWLDSGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR

      orf146-1.pep  FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
15     orf146ng-1    FMLADNLADCSKMIAEISNGRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISP

      orf146-1.pep  AMMEAMQHAHRKIVNTTELLTTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
      orf146ng-1    SMMEAMQHAHRKIVNTTELLTTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAALING

20     orf146-1.pep  RHARRIRIDTAINPELEALA EHLHYQWQGF LWLSTNM RQEISALVILLQRTRRKWLD AHE
      orf146ng-1    RHARRIRIDTAINPELEALA EHLHYQWQGF LWLSTNM RQEISALVILLQRTRRKWLD AHE

      orf146-1.pep  RQHLRQSLLE TREHGX
25     orf146ng-1    RQHLRQSLLE TREHGX

```

Furthermore, ORF146ng-1 (SEQ ID NO: 638) shows homology with a hypothetical *E.coli* protein (SEQ ID NO: 1150):

```

30     sp|P33011|YEEA_ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
      )gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
      Accession Number P33011] [Escherichia coli] )gi|1736682|gnl|PID|d1016560 (D90839)
      ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]
      )gi|1788318 (AE000292) f352; 100% identical to fragment YEEA_ECOLI SW: P33011 but
35     has 203 additional C-terminal residues [Escherichia coli] Length = 352
      Score = 109 bits (271), Expect = 2e-23
      Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

      Query: 20  YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAIYSNAVERML 79
      YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
      Sbjct: 15  YRHYRIVHGT RVALAFLLTFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74

40     Query: 80  GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGKNGYVPMLAGLTMCMMLI 139
      GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++
      Sbjct: 75  GTVLGSILGLIALQLE---LISLPLMLVWCAAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

      Query: 140  GDNGSEWLDSGLMRAMNVLIGXXXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMIAEISN 199
      G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +

```

Sbjct: 132 GSPTGE-IDTALWRS GDVILGSL LAMLF TGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190

Query: 200 GRRMTRERLEQN MVKMRQINAR MVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259  
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V

Sbjct: 191 PNLLERPRLESHLQKLL--TDAVKMRGLIAPASKETRI PKSIYEGIQ TINRNLCMLEL 247

5 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXAALINGRHARRIRIDTAINPEL 316  
+ LN ++R D AL G +N +

Sbjct: 248 QINAYWATRP SHFVLLNAQKLR--DTQHMMQI LLSLVHALYEGNPQPVFANTEKLND AV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNM RQEISALVILLQRTTRK 354

10 Sbjct: 306 EELRQLLN NHDLKV VETPIYGVV LNMETAHQLELLSNLICRALRK 352  
E L + L H+ + G++WL+ ++ L L+ R RK

On the basis of this analysis, including the identification of several transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 15 Example 76

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 639)

20 1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA  
51 GGGCAAAC TC GTCA GTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA  
101 AGATTGT CGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTC CGAT  
151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAAACTCG CCCGCCGCGT  
201 GCGTGAGGCC GGGTTTAAAG TCGTTC CCGT CGTGGGCGCA AC.GCGGTGA  
251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC  
301 GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG  
351 GGTGCGGGCG GCGTTTCCTA TCGTCATGTT TGAAACGCCG CACCGCATCG  
25 401 GTG CAGCGCT TGCCGATATG GCGGAACTGT TCCCCGAACG CCGATTAATG  
451 CTGCGCGCGC AAATTACGAA AACGTTTGAA ACGTTCCTTA GCGGCACGGT  
501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG  
551 AGATGGTGTT GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG  
601 TCCGAGTCCG CGCAAAACAT CATGAAAATC CTCACAGCCG AGCTGCCGAC  
30 651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GGAAAGAAAG  
701 CTTTGTACGA T..

This corresponds to the amino acid sequence (SEQ ID NO: 640; ORF147):

35 1 ..AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD  
51 AGT PAVCDPG AKLARRVREA GFKVVPV VGA XAVMAALSVA GVEGSDFYFN  
101 GFVPPKSGER RKLFAKWRA AFPIVMFETP HRIGAALADM AELFPERRLM  
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL  
201 SESAQNIMKI LTAE LPTKQA AE LAAKITGE GKALYD..

40 Further work revealed the complete nucleotide sequence (SEQ ID NO: 641):

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC  
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATT CAGG GCAAAC TCGT



201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT  
 251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG  
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA  
 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG  
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGC  
 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG  
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA  
 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA  
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG  
 701 TGCTTTATCC GGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG  
 751 CAAAACATCA TGAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
 801 GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC  
 851 TGGCTCTGTC TTGAAAAAC AAATAG

This corresponds to the amino acid sequence (SEQ ID NO: 642; ORF147-1):

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQKLVSVR EHNERQMA DKIVGYLSDGMV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE  
 201 ITKTFFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with hypothetical protein ORF286 (SEQ ID NO: 1151) of *E.coli* (accession number U18997)

ORF147 (SEQ ID NO: 640) and *E.coli* ORF286 protein (SEQ ID NO: 1151) show 36% aa identity in 237aa overlap:

30 Orf147: 1 AEDTRVTAQLLSAYGIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPG 60  
 AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG  
 Orf286: 43 AEDTRHTGLLLQHFGINARLFALHDHNEQQAETLLAKLQEGQNIALVSDAGTPLINDPG 102  
 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFKQWVRA 120  
 L R RE F + GF+P KS RR  
 Orf286: 103 YHLVRTCREAGIRVVPLPGCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAI EAE 162  
 35 Orf147: 121 AFPIVMFETPHRIG AALADMAELFPERR-LMLAREITKTFFETFLSGTVGEIQTALSADGD 179  
 ++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +  
 Orf286: 163 PRTLIFYESTHRLDSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGE LLAWVKEDEN 222  
 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALY 236  
 + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY  
 40 Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAE LPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF147 (SEQ ID NO: 640) shows 96.6% identity over a 237aa overlap with ORF75a (SEQ ID NO: 290) from strain A of *N. meningitidis*:

[illegible]

ORF147a is identical to ORF75a (SEQ ID NO: 290), which includes aa 56-292 of ORF75 (SEQ ID NO: 286).

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF147 (SEQ ID NO: 640) shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) (SEQ ID NO: 644) from *N. gonorrhoeae*:

35	orf147.pep	AEDTRVTAQLLSAYGIQGKLVSVREHNERQ	30
	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGRVSVREHNERQ	85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA	90
	orf147ng	MADKIVGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA	145
40	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGAALADMAELFPERRLM	150
	orf147ng	GVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATLADMAELFPERRLM	205

-453-

```

orf147.pep  LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI  210
            |||||
orf147ng    LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNAMKI  265

orf147.pep  LTAE LPTKQAAELAAKITGEGKKALYD 237
            |:|||||
orf147ng    LAELPTKQAAELAAKITGEGKKALYDLALSWKNK 300

```

An ORF147ng nucleotide sequence (SEQ ID NO: 643) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 644):

```

10      1  MSVFQTAFFM  FQKHLQKASD  SVVGGTLYVV  ATPIGNLADI  TLRALAVLQK
      51  ADIICAEDTR  VTAQLLSAYG  IQGRLVSVRE  HNERQMADKV  IGFLSDGLV
101     101  AQVSDAGTPA  VCDPGAKLAR  RVREAGFKVV  PVVGASAVMA  ALSVAGVAES
151     151  DFYFNGFVPP  KSGERRKLF  KQVRAAFPVV  MFETPHRIGA  TLADMAELFP
201     201  ERRMLLAREI  TKTFTFETLSG  TVGEIQTALA  ADGNQSRGEM  VLVLYPAQDE
251     251  KHEGLSESAQ  NAMKILAAEL  PTKQAAELAA  KITGEGKKAL  YDLALSWKNK
15      301  *

```

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 645):

```

20      1  ATGTTTCAGA  AACACTTGCA  GAAAGCCTCC  GACAGCGTCG  TCGGAGGGAC
      51  ATTATACGTG  GTTGCCACGC  CCATCGGCAA  TTTGGCAGAC  ATTACCCTGC
101     101  GCGCTTTGGC  GGTATTGCAA  AAGGCGGACA  TCATTGTGTC  CGAAGACACG
151     151  CGCGTTACTG  CGCAGCTTTT  GAGCGCGTAC  GGCATTTCAGG  GCAGGTTGGT
201     201  CAGTGTGCGC  GAACACAACG  AGCGGCAGAT  GCGCGACAAG  GTAATCGGTT
251     251  TCCTTTTCAGA  CGGCCTGGTT  GTGGCGCAGG  TTTCCGATGC  GGGTACGCCG
25      301  GCCGTGTGCG  ACCCGGGCGC  GAAACTCGCC  CGCCGCGTGC  GCGAAGCAGG
      351  GTTCAAAGTC  GTTCCCGTCG  TGGGCGCAAG  CGCGGTAATG  GCGGCGTTGA
401     401  GTGTGGCCGG  TGTGGCGGAA  TCCGATTTTT  ATTTCAACGG  TTTGTACCG
451     451  CCGAAATCGG  GCGAACGTAG  GAAATTGTTT  GCCAAATGGG  TGCGGGCGGC
501     501  ATTTCTGTGC  GTCATGTTTG  AAACGCCGCA  CCGAATCGGG  GCAACGCTTG
551     551  CCGATATGGC  GGAATTGTTT  CCCGAACGCC  GTCTGATGCT  GGCAGCGGAA
30      601  ATCACGAAAA  CGTTTGAAAC  GTTCTTAAGC  GGCACGGTTG  GGGAAATTCA
651     651  GACGGCATTG  GCGGCGGACG  GCAACCAATC  GCGCGGCGAG  ATGGTGTTGG
701     701  TGCTTTATCC  GGCGCAGGAT  GAAAAACACG  AAGGCTTGTC  CGAGTCTGCG
751     751  CAAAATGCGA  TGAATACTCT  TGCGGCCGAG  CTGCCGACCA  AGCAGGCGGC
801     801  GGAGCTTGCC  GCCAAGATTA  CAGGTGAGGG  CAAAAGGCT  TTGTACGATT
35      851  TGGCACTGTC  GTGGAATAAC  AAATGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 646; ORF147ng-1):

```

40      1  MFQKHLQKAS  DSVVGGTLYV  VATPIGNLAD  ITLRALAVLQ  KADIICAEDT
      51  RVTAQLLSAY  GIQRLVSVR  EHNERQMADK  VIGFLSDGLV  VAQVSDAGTP
101     101  AVCDPGAKLA  RRVREAGFKV  VPVGASAVM  AALSVAGVAE  SDFYFNGFVP
151     151  PKSGERRKLF  AKWVRAAFPV  VMFETPHRIG  ATLADMAELF  PERRMLLARE
201     201  ITKTFTFELS  GTVGEIQTAL  AADGNQSRGE  MVLVLYPAQD  EKHEGLSESA
251     251  QNAMKILAAE  LPTKQAAELA  AKITGEGKKA  LYDLALSWKN  K*

```

45 ORF147ng-1 (SEQ ID NO: 646) shows homology to a hypothetical *E.coli* protein (SEQ ID NO: 1152):

```

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
(F286)
)gi|606086 (U18997) ORF_f286 [Escherichia coli]

```

)gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region  
 [Escherichia coli] Length = 286  
 Score = 218 bits (550), Expect = 3e-56  
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

5 Query: 4 KHLQKASDSVVGGLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63  
 K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI  
 Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFGIN 59

10 Query: 64 GRLVSVREHNERQMADKVI GFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123  
 RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+  
 Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFQKVVRAAFPVVMFETPHRIGATL 183  
 G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L  
 Sbjct: 120 PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAI EAEPRTLIFYESTHRLDLSL 179

15 Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242  
 D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +  
 Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

Query: 243 HEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLAL 286  
 E L A + +L AELP K+AA LAA+I G K ALY AL  
 Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKNALYKYAL 282

20

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 77

25 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 647)

30 1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA  
 51 AACCGGTGCG ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCTG  
 101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC  
 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG  
 201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT  
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC  
 301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG  
 351 GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT  
 401 CAACAwCGww TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG  
 35 451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCATA  
 501 AATWTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG  
 551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC  
 601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCGCGAAA  
 651 GTTCATATCA TATTGCAAGT .....  
 40 701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA  
 751 AAAGTGGTTA ATTAATGGGG TATTGCAAAC GGGCAACCCC TATATAGGAA  
 801 AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC  
 851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACGTC AAAATGGGAA  
 901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC  
 45 951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG  
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC  
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT

-455-

5 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT  
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA  
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA  
 1251 CCGTTACTTG GAAAGTAAAC GGCGTGGCAA ACGACCGCCT GTCCAAAATC  
 1301 GGCAAAGGCA CGCTG.....  
 //

10 2101 ..... GATAAAG  
 2151 TGA CTGCTT ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC  
 2201 GATCAGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT  
 2251 TAGTCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA  
 2301 ACGGCAACCK TagCctCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC  
 2351 ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG  
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCGGC AACGCTAAGG  
 2451 CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG  
 15 2501 GCAGTATTCC ATTTTGAAAG CAGCCGCTTT ACCGGACAAA TCAGCGGCGG  
 2551 CAagGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCag  
 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT  
 2651 TCCGCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA  
 2701 TGCGCCGCGC CGCGGTTGCG GCCGTTGCGC CCGTTCCTTA TTATmCGTTA  
 20 2751 CACCGCCAAC TTCGTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC  
 2801 AAATTGAACG GTCAGGGAAC ATTCGCTTT ATGTCGGAAC TCTTCGGCTA  
 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT  
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG  
 25 2951 GTAGTGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC  
 3001 CCTGCAAAAC GAACACGTG ATGCAGGCGC GTGG.....  
 //

30 3551 ..... TTAGAC CGCGTATTTG CCGAAGACCG  
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT  
 3651 CGCAAGATTT CCGCGCTAC CGCCAACAAA CCGACCTGCG CCAAATCGGT  
 3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGATCCTGT TTTGCGACAA  
 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG  
 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC  
 3851 ATCAGuCGCG GGCGCGGGTT TTAGCAGCGG CAGCCTTTcA GACGGCATCG  
 35 3901 GAGsmAAAwT CCGCCCGCGC GTGctGCATT ACGGCATTCA GGCACGAtAC  
 3951 CGCGCCGgtt tCggCGgAtT CGGCATCGAA CCGCACATCG GCGCAACGCG  
 4001 ctATTTGTC CAAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA  
 4051 CCCCCGCCT TGCATTCAAC CGcTACCGCG CGGCATTAA GGCAGATTAT  
 4101 TCATTCAAAC CGGCGCAACA CATTTCCATC ACGCCTTATT TGAGCCTGTC  
 40 4151 CTATACCGAT GCCGTTTCGG GCAAAGTCCG AACACGCGTC AATACCGCCG  
 4201 TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC  
 4251 GCCGAAATCA AAGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG  
 4301 CCCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGCTACCGCT  
 4351 GGTAA...

45 This corresponds to the amino acid sequence (SEQ ID NO: 648; ORF1):

50 1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFaEN KGKFAVGAKD IEVYNKKGEL VGKSM TKAPM IDFSVVS RNK  
 101 VAALVG VQYI VSVAHNGGYN NVDFGAEGXN IXDQXRXYK IVKRNNYKAG  
 151 TKGHPY GGDY HMPRLHXKVT DAEPVENTSY MDGRKYIDQN NYPDVRVIGA  
 201 GRQYWRSDED EPNNRESSYH IAS.....GS PMFIYDAQKQ  
 251 KWLINGVLQT GNPYIGKSNG FQLVRKDWFY DEIFAGDTHS VFYEPRQNGK  
 301 YSFNDNNGT GKINAKHEHN SLPNRLKTRT VQLFNVSLSE TAREPVYHAA  
 351 GGVNSYRPRL NNGENISFID EGKGELILTS NINQGAGGLY FQGDFTVSPE  
 401 NNETWQGAGV HISEDSTVTW KVNGVANDRL SKIGKGTL.....  
 55 //

701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL  
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS  
 801 DHAVQNGSLT LSGNAKANVS HSALNGNVSL ADKAVFHFES SRFTGQISGG  
 851 KDTALHLKDS EWTLP SGXEL GNLNLDNATI TLNSAYRHDA AGAQTGSATD

901 APRRRSRRSR RSLXVTPPT SVESRFNTLT VNGKLNQGT FRFMSELFY  
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLNFT  
 1001 LQNEHVDAGA W.....  
 5 1151 .....  
 1201 RNAVWTSGIR DTKHYRSQDF RAYRQOTDLR QIGMQKNLGS GRVGILFSHN  
 1251 RTENTFDDGI GNSARLAHGA VFGQYGIDRF YIGISAGAGF SSGSLSDGIG  
 1301 XKXRRRVLHY GIQARYRAGF GFGIEPHIG ATRYFVQKAD YRYENVNIAT  
 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAAAG KVRTRVNTAV  
 10 1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW  
 1451 \*

Further sequencing analysis revealed the complete nucleotide sequence (SEQ ID NO: 649):

1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA  
 51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCGT  
 15 101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACCTATTT CGGCATCAAC  
 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG  
 201 GCGCAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT  
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC  
 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG  
 20 351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC  
 401 ATCGTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT  
 451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCATAAAAT  
 501 TGTCACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA  
 551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCGTAT TGGGGCAGGC  
 25 601 AGGCAATATT GGCGATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTC  
 651 ATATCATATT GCAAGTGCGT ATTCTTGGCT CGTTGGTGGC AATACCTTTG  
 701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAAAAAATT  
 751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG  
 801 TGGCTACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA  
 30 851 ATGGGGTATT GCAAACGGGC AACCCTTATA TAGGAAAAAG CAATGGCTTC  
 901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC  
 951 CCATTAGTA TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAACG  
 1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT  
 1051 CTGCCTAATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT  
 35 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCACAA  
 1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCCTT TATTGACGAA  
 1201 GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG  
 1251 ATTATATTTC CAAGGAGATT TTACGGTCTC GCCTGAAAAT AACGAACTT  
 1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA  
 40 1351 GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT  
 1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG  
 1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC  
 1501 TTTAGTGAAA TCGGCTTGGT CAGCGCGAGG GGTACGGTGC AACTGAATGC  
 1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC  
 45 1601 GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC  
 1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT  
 1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACAAACGCT  
 1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAGAT  
 1801 ACGACCAAAA CGAACGGCG GCTCAACCTT GTTTACCAGC CCGCGCAGAG  
 50 1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA  
 1901 CGCAAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC  
 1951 TACAATCATT TAAACGACCA TTGGTCGCAA AAAGAGGGCA TTCCTCGCGG  
 2001 GGAAATCGTG TGGGACAAAC ACTGGATCAA CCGCACATTT AAAGCGGAAA  
 2051 ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCTCGCA TGTTCGCAAA  
 55 2101 GTGAAAGGCG ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTGGTGT  
 2151 CGCACCGCAT CAAACGCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC  
 2201 TGACAAATTG TGTGCAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA  
 2251 TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGCCG ATCAGCTCA  
 2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAAATG

5 2351 GCGATACACG TTATACAGTC AGCCACAACG CCACCCAAAA CGGCAACCTT  
 2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG  
 2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG  
 2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC  
 2551 CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTTCCA  
 2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG  
 2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAATTA  
 2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCCTATCG  
 10 2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TGCACAGAT GCGCCGCGCC  
 2801 GCCGTTTCGCG CCGTTCGCGC CGTTCCTTAT TATCCGTTAC ACCGCCAACT  
 2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG  
 2901 TCAGGGAACA TTCCGCTTTA TGTGGAAGT CTTCCGCTAC CGCAGCGACA  
 2951 AATTGAAGCT GCGGAAAGT TCCGAAGGCA CTTACACCTT GCGCGTCAAC  
 15 3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGAAGG  
 3051 AAAAGACAAAC AAACCGCTGT CCGAAAACCT TAATTTACC CTGCAAAACG  
 3101 AACACGTCGA TGCCGGCGCG TGGCGTTACC AACTCATCCG CAAAGACGGC  
 3151 GAGTTCCGCC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT  
 3201 CGGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAAC GCGCAAAGCC  
 3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC  
 20 3301 GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCG GCATTATGCA  
 3351 GGCGGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG  
 3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC  
 3451 GCGCGCCGCG CCGCCCGGGA TTTGCCGCAA CTGCAACCCC AACCAGCAGC  
 25 3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAAATAGC GGTTCGAGTG  
 3551 AATTTTCCGC CACGCTCAAC AGCGTTTTTCG CCGTACAGGA CGAATTAGAC  
 3601 CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG  
 3651 GGACACCAAA CACTACCGTT CGCAAGATTT CCGCGCCTAC CGCCAACAAA  
 3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC  
 3751 GGCATCCTGT TTTGCGACAA CCGGACCGAA AACACCTTCG ACGACGCAT  
 30 3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTTCGGG CAATACGGCA  
 3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC  
 3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCCGCG TGCTGCATTA  
 3951 CGGCATTACG GCACGATACC GCGCCGGTTT CCGCGGATTC GGCATCGAAC  
 35 4001 CGCACATCGG CGCAACGCGC TATTTTCGTCC AAAAAGCGGA TTACCGCTAC  
 4051 GAAAACGTCA ATATCGCCAC CCCCGGCCTT GCATTCAACC GCTACCGCGC  
 4101 GGGCATTTAAG GCAGATTATT CATTCAAACC GCGCAACAC ATTTCCATCA  
 4151 CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA  
 4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTTCGGCA AAACCCGCAG  
 4251 TCGGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC  
 40 4301 ACGTGCCGCG CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC  
 4351 ATCAAATTAG GCTACCGCTG GTAA

This corresponds to the amino acid sequence (SEQ ID NO: 650; ORF1-1):

45 1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFEEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG  
 101 VAALVGDQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT  
 151 KGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG  
 201 RQYWRSEDEE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI  
 251 KHSPYGFLLP GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNFG  
 50 301 QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS  
 351 LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE  
 401 GKGLILTSN INQGAGGLYF QGDFTVSPEN NETWQGAGVH ISEDSTVTWK  
 451 VNGVANDRLS KIGKGLTHVQ AKGENQGSIS VGDGTVILDQ QADDKGGKQA  
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIONT  
 55 551 DEGAMIVNHN QDKESTVTIT GNKDIAATGN NNSLDSKKEI AYNGWFGKEK  
 601 TTKTNGRLNL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA  
 651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK  
 701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS  
 751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL

5

10

15 Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF1 (SEQ ID NO: 648) shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) (SEQ ID NO: 652) from strain A of *N. meningitidis*:

20

25

30

35

40

45



5

```

      330      340      350      360      370      380
orfl.pep  SLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLYFQGDFT
orfla     ||:|| :||| |||||:|||||:||| ||:|||:|||||:|||
      360      370      380      390      400      410
orfla     SLNETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYNGNKLILSNINQGAGGLYFEGDFT

```

10

```

      390      400      410      420      430
orfl.pep  VSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGTL-----
orfla     ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      420      430      440      450      460      470
orfla     VSPENNETWQGAGVHISEDSTVTWKVNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGT

```

15

```

orfl.pep  -----
orfla     VILDQQADDKGGKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRDLNNGHSLSFH
      480      490      500      510      520      530

```

20

```

orfl.pep  -----
orfla     RIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFGKEDTTK
      540      550      560      570      580      590

```

25

```

orfl.pep  -----
orfla     TNGRLNLVYQPAAEDRTXLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSGWSKMEG
      600      610      620      630      640      650

```

30

```

orfl.pep  -----
orfla     IPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKEGDXHLSNHAQAVFGVAPHQSH
      660      670      680      690      700      710

```

35

```

      440      450      460      470      480
orfl.pep  -----XXXXXDKVTASLTKTDISGNVDLADHAHLNLTGLATLNGNLSAN
      : || : || ||||| || || |||||
orfla     TICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXXXLXGXAXLXGNLSAN
      720      730      740      750      760      770

```

40

```

      490      500      510      520      530      540
orfl.pep  GDTRYTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTSLG
orfla     ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      780      790      800      810      820      830
orfla     GDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNGSLTSLD

```

45

```

      550      560      570      580      590      600
orfl.pep  NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGNL
orfla     ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      840      850      860      870      880      890
orfla     NAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSGTGLGNL

```

50

```

      610      620      630      640      650      660
orfl.pep  NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRLXVTPPTSVESRFNTLTVNG
orfla     ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      900      910      920      930      940      950
orfla     NLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFNTLTVNG

```

5

The complete length ORF1a nucleotide sequence (SEQ ID NO: 651) is:

1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA  
51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCGT

5 101 TCGGCATTCT TCCCCAAGCT TGGGCGGGAC ACACTTATTT CGGCATCAAC  
 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCCG  
 201 GGCGAAAAGAT ATTGAGGTNT ACAACAAAAA AGGGGAGTTG GTCCGCAAAAT  
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC  
 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG  
 351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAAGNAAT CCCGATCAGC  
 401 ACCGTTTTTC TTACCAAATT GTGAAAAGAA ATAATTATAA GCCTGACAAAT  
 451 TCACACCCCTT ACAACGGCGA TTANCATATG CCGCGTTTGC ATAAATTTGT  
 10 501 CACAGATGCA GAACCTGTCT AAATGACGAG TGACATGAGG GGGAAATACCT  
 551 ATTCGATATA AGAAAAATAT CCCGAGCGTG TCCGCATCGG CTCAGGACAC  
 601 CACTATTGGC GTTATGATGA TGACAAACAC GCGGATTTAT CCTACTCCGG  
 651 CGCATGGTTA ATTGGCGGCA ATACACATAT GCAGGGTTGG GGAAATAATG  
 701 GCGTANTTAG TTTGAGCGGC GATGTGCGCC ATGCCAACGA CTATGGCCCT  
 751 ATGCCGATTG CAGGTGCGGC AGGCGACAGC GGTTCGCCAA TGTTTATTTA  
 15 801 TGACAAAACA AACAATAAAT GGCTGCTCAA CGGAGTTTAA CAAACCGGCT  
 851 ACCCTTATTC CCGCAGGGAA AACGGTTTCC AGCTGATACG CAAAGATTGG  
 901 TTCTACGATG ACATTTACAG AGGCGATACA CATACCGTCT NTTTTGAACC  
 951 GCGCAGTAAC GGACATTTTT CTTTTACATC CAACAACAAC GGTACGGGTA  
 20 1001 CGGTAACAGA AACCAACGAA AAGGTNTCCA ATCCAAAGCT TAAAGTACAG  
 1051 ACAGTCCGAC TGTTTGACGA ATCTTTGAAT GAAACTGATA AAGAACCAGT  
 1101 TTACGCGGCA GGGGGTGTGA ATCAGTACCG TCCAAGGTTA AACAACGGTG  
 1151 AAAACCTTTC TTTTATCGAT TACGGCAACG GCAAACCTCAT CTTATCAAAC  
 1201 AACATCAACC AAGGCGCGGG CGGTTTGTAT TTTGAAGGTG ATTTTACGGT  
 1251 CTCGCCTGAA AACAACGAAA CGTGGCAAGG CGCGGGCGTT CATATCAGTG  
 25 1301 AAGACAGTAC CGTTACTTGG AAAGTAAACG GCGTGGCAAA CGACCGCCTG  
 1351 TCCAAAATCG GCAAAGGCAC GCTGCACGTT CAAGCCAAAG GGGAAAACCA  
 1401 AGGCTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGCGAG  
 1451 ACGATAAAGG CAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGGC  
 30 1501 AGGGGTACGG TGCAACTGAA TGCCGATAAT CAGTTCAACC CCGACAAACT  
 1551 CTATTTTCGGC TTTTCGCGCG GACGTTTGA TTTAAACGGG CATTCGCTTT  
 1601 CGTTCCACCG TATTCAAAAT ACCGATGAAG GGGCGATGAT TGNCNATCAT  
 1651 AATGCCACAA CAACATCCAC CGTTACCATT ACAGGGAATG AAAGTATTAC  
 1701 ACAACCGAGT GGTAAGAAAT TCAATAGACT TAATTACAGC AAAGAAATTG  
 1751 CCTACAACGG TTGGTTTGGC GAGAAAGATA CGACCAAAAC GAACGGGCGG  
 35 1801 CTCAACCTTG TTTACCAGCC CGCCGCAGAA GACCGCACCC NGCTGCTTTC  
 1851 CGGCGGAACA AATTTAAACG GCAACATCAC GCAAACAAAC GGCAAACTGT  
 1901 TTTTCAGCGG CAGACCGACA CCGCACGCCT ACAATCATTT AGGAAGCGGG  
 1951 TGGTCAAAAA TGGAAGGTAT CCCACAAGGA GAAATCGTGT GGGACAACGA  
 40 2001 CTGGATCNAC CGCACGTTTA AAGCGGAAAA TTTCCATATT CAGGGCGGGC  
 2051 AGGCGGTGAT TTCCCGCAAT GTTGCCAAAG TGGAAGGCGA TTGNCAATTG  
 2101 AGCAATCACG CCCAAGCAGT TTTTGGTGTC GCACCGCATC AAAGCCATAC  
 2151 AATCTGTACA CGTTCGACT GGACNGTCT GACAAATTGT GTCGAANAAA  
 2201 NCATTACCGA CGATAAAGTG ATTGCTTCAT TGAATAAGAC NGACNTNAGC  
 2251 GGCANTGTNA GNCTNNCCNA TNACGNTNNT TNAAANCTCN CNGGGCNTGC  
 45 2301 NNCATNAAN GGCAATCTTA GTGCAAATGG CGATACACGT TATACAGTCA  
 2351 GCCACAACGC CACCCAAAAC GGCAACCTTA GCCTCGTGGG CAATGCCCAA  
 2401 GCAACATTTA ATCAAGCCAC ATTAACGGC AACNCATCGG NTTCCGGCAA  
 2451 TGCTTCATTT AATCTAAGCA ACAACGCCGC ACAAACGGC AGTCTGACGC  
 50 2501 TTTCCGACAA CGCTAAGGCA AACGTAAGCC ATTCCGCACT CAACGGCAAT  
 2551 GTCTCCCTAG CCGATAAGGC AGTATTCCAT TTTGAAAACA GCCGCTTTAC  
 2601 CGGACAACCT AGCGGCAGCA AGGANACAGC ATTACACTTA AAAGACAGCG  
 2651 AATGGACGCT GCCGTACGGC ACGGAATTAG GCAATTTAAA CCTTGACAAC  
 2701 GCCACCATTA CACTCAATTC CGCCTATCGC CACGATGTGT CAGGCGCGCA  
 2751 AACCGGCAGN GTGTACAGCA CGCCGCGCCG CCGTTCGCGC CGTTCCTAT  
 55 2801 TATCCGTTAC ACCGCCAAT TCGGTAGAAT CCCGTTTCAA CACGCTGACG  
 2851 GTAAACGGCA AATTGAACNG TCAAGGAACA TTCCGCTTTA TGTCCGAACT  
 2901 CTTCCGGTAC CGAAGCGACA AATTGAAGCT GGCGGAAAGT TCCGAAGGNA  
 2951 CTTACACCTT GGCGGTCAAC AATACGGGCA ACGAACCCGT AAGCCTCGAT  
 3001 CAATTGACGG TAGTGGAAGG GAAAGACAAC AAACCGCTGT CCGAAAACCT  
 60 3051 TAATTTACC CTGCAAAACG AACACGTCGA TGCCGGCGCG TGGCGTTACC  
 3101 AACTCATCCG CAAAGACGGC GAGTTCGCC TGCATAATCC GGTCAAAGAA

5 3151 CAAGAGCTTT CCGACAAACT CGGCAAGGCA GAAGCCAAAA AACAGGCCGA  
 3201 AAAAGACAAAC GCGCAAAGCC TTGACGCGCT GATTGCGGCC GGGCGCGATG  
 3251 CCGCCGAAAA GACAGAAAGC GTTGCCGAAC CGGCCCGGCN GGCAGGCGGG  
 3301 GAAAATGTCG GCATTATGCA GGCGGAGGAA GAGAAAAAAC GGGTGCAGGC  
 3351 GGATAAAGAC AGCGCNTTGG CGAAACAGCG CGAAGCGGAA ACCCGCCCGG  
 3401 NTACCACCGC CTTCCCCCGC GCCCCGNGCG CCCGCCGGGA TTTGCCGCAA  
 3451 CCGCAGCCCC AACCGCAACC TCAACCCCAA CCGCAGCGCG ACCTGATNAG  
 3501 CCGTTATGCC AATAGCGGTT TGAGTGAATT TTCCGCCACG CTCAACAGCG  
 10 3551 TTTTCGCCGT ACAGGACGAA TTGGACCGCG TGTTCGCCGA AGACCGCCGC  
 3601 AACGCNGTTT GGACAAGCNG CATCCGNAC ACCAAACACT ACCGTTTCGCA  
 3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC  
 3701 AGAAAAACCT CGGCAGCGGG CGCGTCGCA TCCTGTTTC GCACAACCGG  
 3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA  
 3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTTCGAC ATCGGCATCA  
 15 3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC  
 3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC  
 3951 CGGTTTCGGC GGATTTCGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT  
 4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC  
 4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCATT  
 20 4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCTTATA  
 4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG  
 4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGGC GAATGGGGCG TAAACGCCGA  
 4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGNCCGC  
 25 4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 652):

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG  
 101 VAALVGDAQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNRYKPDN  
 30 151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDKEY PERVRIGSGH  
 201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP  
 251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGVL QTGYPPYSGRE NGFQLIRKDW  
 301 FYDDIYRGDT HTVXFEPNRSN GHFSFTSNMN GTGTVTETNE KVSNNPKLVQ  
 351 TVRLFDESIN ETDKEPVYAA GGVNQYRPRL NNGENLSFID YGNGKLILSN  
 35 401 NINQGAGGLY FEGDFTVSP E NNETWQAGV HISEDSTVTW KVNQVANDRL  
 451 SKIGKGTLLHV QAKGENQSGI SVGDGTVILD QQADDKGGKQ AFSEIGLXSG  
 501 RGTVQLNADN QFNPKLYFG FRGGRDLNG HSLSFHRIQN TDEGAMIXXH  
 551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR  
 601 LNLVYQPAAE DRTXLLSGGT NLNGNITQTN GKLFSSGRPT PHAYNHLGSG  
 40 651 WSKMEGIPQG EIVWDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGDHL  
 701 SNHAQAVFGV APHQSHICT RSDWTGLTNC VEXXITDDKV IASLTKTDXS  
 751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ  
 801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN  
 851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNLNDN  
 45 901 ATITLNSAYR HDAAGAQTGX VSDTPRRRSR RSLLSVTPPT SVESRNTLT  
 951 VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD  
 1001 QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE  
 1051 QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAAEKTES VAEPARXAGG  
 1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDLPO  
 50 1151 PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR  
 1201 NAVWTSXIRX TKHYRSQDFR AYRQQTDLRQ IGMQKNLGS RVGILFSHNR  
 1251 TENXFDDGIG NSARLAHGA FGQYIGIRFD IGISTGAGFS SGXLSDGIGG  
 1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIATP  
 1351 GLAFNRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL  
 55 1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAAAA KGPQLEAQHS AGIKLGYRW\*

A transmembrane region is underlined.

ORF1-1 (SEQ ID NO: 650) shows 86.3% identity over a 1462aa overlap with ORF1a (SEQ ID NO: 652):

5	orf1a.pep	10 20 30 40 50 60	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN
	orf1-1	10 20 30 40 50 60	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN
10	orf1a.pep	70 80 90 100 110 120	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN
	orf1-1	70 80 90 100 110 120	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN
15	orf1a.pep	130 140 150 160 170 179	NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAPVEMTSDM
	orf1-1	130 140 150 160 170 180	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAPVEMTSYM
20	orf1a.pep	180 190 200 210 220 230	RGNTYSDKEKYPERVIRIGSGHHYWRYYYYDKHGDL--SYSGA---WLIGGNTHMQGWGNN
	orf1-1	190 200 210 220 230 240	DGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG
25	orf1a.pep	240 250 260 270 280 290	GVXSLSGD-VRHANDYGPMPPIAGAAGDSGSPMFIYDKTNNKWLNLGVLQTGYPYSGRENG
	orf1-1	250 260 270 280 290	GTVNLGSEKIKHS-PYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG
30	orf1a.pep	300 310 320 330 340 350	FQLIRKDWFYDDIYRGDTHTVXFEPRSNGHFSFTSNNGTGTVTETNEKVSNP-KLKVQT
	orf1-1	300 310 320 330 340 350	FQLVRKDWFYDEIFAGDTHSVFYEPQRNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT
35	orf1a.pep	360 370 380 390 400 410	VRLFDESNETDKEPVY-AAGGVNQYRPLNNGENLSFIDYGNGKLILSNINQAGAGGLY
	orf1-1	360 370 380 390 400 410	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQAGAGGLY
40	orf1a.pep	420 430 440 450 460 470	FEGDFTVSPENNETWQAGAGVHISEDSTVTWKVNGVANDRLSKIGKGLHVLQVAKGENQGS
	orf1-1	420 430 440 450 460 470	FQGDFTVSPENNETWQAGAGVHISEDSTVTWKVNGVANDRLSKIGKGLHVLQVAKGENQGS
45	orf1a.pep	480 490 500 510 520 530	SVGDGTVILDQQADDKGKKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRDLNLG
	orf1-1	480 490 500 510 520 530	SVGDGTVILDQQADDKGKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNLG

5	orfla.pep	540	550	560	570	580	590
	orfl-1	540	550	560	570	580	590
10	orfla.pep	600	610	620	630	640	650
	orfl-1	600	610	620	630	640	650
15	orfla.pep	660	670	680	690	700	710
	orfl-1	660	670	680	690	700	710
20	orfla.pep	720	730	740	750	760	770
	orfl-1	720	730	740	750	760	770
25	orfla.pep	780	790	800	810	820	830
	orfl-1	780	790	800	810	820	830
30	orfla.pep	840	850	860	870	880	890
	orfl-1	840	850	860	870	880	890
35	orfla.pep	900	910	920	930	940	
	orfl-1	900	910	920	930	940	950
40	orfla.pep	950	960	970	980	990	1000
	orfl-1	960	970	980	990	1000	1010
45	orfla.pep	1010	1020	1030	1040	1050	1060
	orfl-1	1020	1030	1040	1050	1060	1070
50	orfla.pep	1070	1080	1090	1100	1110	1120
	orfl-1	1080	1090	1100	1110	1120	1130

5	orf1a.pep	1130	1140	1150	1160	1170	1180
	orf1-1	1140	1150	1160	1170	1180	1190
10	orf1a.pep	1190	1200	1210	1220	1230	1240
	orf1-1	1200	1210	1220	1230	1240	1250
15	orf1a.pep	1250	1260	1270	1280	1290	1300
	orf1-1	1260	1270	1280	1290	1300	1310
20	orf1a.pep	1310	1320	1330	1340	1350	1360
	orf1-1	1320	1330	1340	1350	1360	1370
25	orf1a.pep	1370	1380	1390	1400	1410	1420
	orf1-1	1380	1390	1400	1410	1420	1430
30	orf1a.pep	1430	1440	1450			
	orf1-1	1440	1450				

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387) (SEQ ID NO: 1153)

Amino acids 23-423 of ORF1 (SEQ ID NO: 648) show 59% aa identity with hap protein (SEQ ID NO: 1153) in 450aa overlap:

35	orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYYRDFAENKGKFAVGAKDIEVYNKKGELVG	82
	hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYYRDFAENKGKFTVGAKNIEVYNKEGQLVG	65
40	orf1	83	KSMTKAPMIDFSVVSRRNGVAALVGVQYIVSVAHNGGYNNVDFGAEGXNIXDQXRXTYKIV	142
	hap	66	TSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGGYNDVDFGAEGRN-PDQHRFTYQIV	124
	orf1	143	KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSYMDGRKYIDQNNYPDRVRIGAGR	202
	hap	125	KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEPVGMTTNMDGKVYADRENYPERVIRIGSGR	184

5

```

orf1 203 QYWRSEDEPNNRESSYHIA----- 222
      QYWR+D+DE N SSY+++
hap 185 QYWRTDKDEETNVHSSYYVSGAYRYLTAGNHTQSGNGNGTVNLSGNVVSPNHYGPLPTG 244

orf1 223 -----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDFYDEIFAGDTHSVF 277
      SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF
hap 245 GSKGDSGSPMFIYDAKKKQWLINAVLQTGHPPFFGRNGFQLIREEFYNEVLAVDTSPSVF 304

orf1 278 --YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334
      Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A
hap 305 QRYIPPINGHYSFVSNNDGTGKLTLTRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV- KA 363

10 orf1 335 AGGVNSYRPRNLNNGENISFIDEGKGELILTSNINQGAGGLYFQGDFTV-SPENNETWQGA 393
      A G N Y+PR+ G+NI D+GKG L + +NINQGAGGLYF+G+F V +NN TWQGA
hap 364 AAGYNIYQPRMEYGKNIYLGDOGKGTLTIENNINQGAGGLYFEGNFVVGKQNNITWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGTL 423
      GV I +D+TV WKV+ NDRLSKIG GTL
15 hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453

```

Amino acids 715-1011 of ORF1 (SEQ ID NO: 648) show 50% aa identity with hap protein (SEQ ID NO: 1153) in 258aa overlap:

20

```

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSHAVQNGSLTSL 98
      DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS
hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTLIDHSQFTLSNNATQTGNIKLS 792

orf1 99 GNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158
      +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N
25 hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSDTLQN 852

orf1 159 LNLDNATITLNSAYRHDAAAGATGSATDAPXXXXXXXXXXLLXVTPPTSVESRFNTLTVN 218
      L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN
hap 853 LTLNNSTVTLNSAY-----SASSNNAPRHRS-----LETETPTPTSAEHRFNTLTVN 899

orf1 219 GKLNQGQTFRFMSELFYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDNKP 278
      GKL+GQGTF+F S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP
30 hap 900 GKLSGQGTFOFTSSLFYKSDKLKLSNDAEGDYTLVRNTGKEPVTLEQLTLIESLDNKP 959

orf1 279 LSENLNFTLQNEHVDAGA 296
      LS+ L FTL+N+HVDAGA
35 hap 960 LSDKLKFTLENDHVDAGA 977

```

Amino acids 1192-1450 of ORF1 (SEQ ID NO: 648) show 41% aa identity with hap protein (SEQ ID NO: 1153) in 259aa overlap:

40

```

Orf1 1 LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR 60
      LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R
hap 1135 LDRLFVDQAQSAVWNTNIAQDKRRYDSADFAYQQKTNLQIGVQKALANGRIGAVFSHSR 1194

orf1 61 TENTFDDGIGNSARLAHGAVFGQYGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120
      ++NTFD+ + N A L + F QY K R+ ++YG
hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGDLOFGVNVGTGISASKMAEEQSRKIHRKAINYG 1254

```



5

orf1	121	IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA	180
		+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P	
hap	1255	VNASYQFRLGQLGIQPYFGVNRIFYERENYQSEEVVRKTPSLAFNRYNAGIRVDYTFPT	1314
orf1	181	QHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA	240
		+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +	
hap	1315	DNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKS	1374
orf1	241	KGPQLEAQHSAGIKLGYRW	259
		+G QL Q + G+KLGYSW	
hap	1375	QGSQLGKQQNVGVKLGYSW	1393

# 10 Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 (SEQ ID NO: 648) show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) (SEQ ID NO: 654) from *N.gonorrhoeae*:

15	orf1.pep	MKTDDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTDDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
20	orf1.pep	NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKKVTDAPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKKFVTDAPVEMTSY	179
	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNRESSYHIAS-----	223
25	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNRESSYHIASAYSWLVGGNTFAQNGSG	239
	orf1.pep	-----GSPMFIYDAQKQKWLINGVLQTNPNYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLLPTGGSGSPMFIYDAQKQKWLINGVLQTNPNYIGKSNG	289
30	orf1.pep	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	FQLVRKDWFYDEIFAGDTHSVFYEPHONGKYFFNDNNGAGKIDAKHKHSLPYRLKTRT	359
	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQAGGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDKGGELILTSNINQAGGGLY	
35	orf1.pep	FQGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
	orf1.pep	// DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
40	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774

	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLS DHA	803
	orf1.ng	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
5	orf1.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1.ng	VQNGSLTSLSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893
	orf1.pep	LPSGXELGNLNLNDNATITLNSAYRHDAAQAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
	orf1.ng	LPSGTELGNLNLNDNATITLNSAYRHDAAQAQTGSAADAPRRRSRRS---LLSVTPPTS AE	950
10	orf1.pep	SRFNTLTVNGKLNQGQTFRFMSELFYRSDKCLKLAESSEGTYTLAVNNTGNEPASLEQLT	983
	orf1.ng	SRFNTLTVNGKLNQGQTFRFMSELFYRSGKCLKLAESSEGTYTLAVNNTGNEPVSLEQLT	1010
	orf1.pep	VVEGKDNKPLSENLNFTLQNEHVDAGAW	1011
	orf1.ng	VVEGKDNTPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAGET	1070
15	orf1.pep	//	
	orf1.ng	LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1.ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
20	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFY	1271
	orf1.ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYGIGRFD	1299
	orf1.pep	IGISAGAGFSSGSLSDGIGKXRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1331
	orf1.ng	IGISAGAGFSSGSLSDGIRGKIRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY	1359
25	orf1.pep	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL	1391
	orf1.ng	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVL	1419
	orf1.pep	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKGPQLEAQHSAGIKLGYRW	1440
	orf1.ng	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKGPQLEAQHSAGIKLGYRW	1468

The complete length ORF1 ng nucleotide sequence was identified (SEQ ID NO: 653):

35	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCTAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCCGCTTA	CTTAGCCATA	TGCCTGTCGT
	101	TCGGCATTCT	GCCCCAAGCC	CGGGCGGGAC	ACACTTATTT	CGGCATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAAT
	251	CGATGACGAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTATC	GCGTAACGGC
40	301	GTGGCGGCAT	TGGCGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGTATAAC	AATGTTGATT	TTGGTGCGGA	GGGAAGCAAT	CCCATCAGC
	401	ACCGCTTTTC	TTACCAAAAT	GTGAAAAGAA	ATAATTATAA	AGCAGGGACT
	451	AACGGCCATC	CTTATGGCGG	CGATTATCAT	ATGCCGCGTT	TGCACAAATT
	501	TGTCACAGAT	GCAGAACCTG	TTGAGATGAC	CAGTTATATG	GATGGGTGGA
	551	AATACGCTGA	TTTAAATAAA	TACCCTGATC	GTGTTCTGAAT	CGGAGCAGGC
45	601	AGACAATATT	GGCGGTCTGA	TGAAGACGAA	CCCAATAACC	GCGAAAGTTC
	651	ATATCATATT	GCAAGCGCAT	ATTCTTGGCT	CGTCGGTGGC	AATACCTTTC
	701	CACAAAATGG	ATCAGGTGGT	GGCACAGTCA	ACTTAGGTAG	CGAAAAAATT
	751	AAACATAGCC	CATATGGTTT	TTTACCAACA	GGAGGCTCAT	TTGGCGACAG

5 801 TGGCTACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA  
851 ATGGGGTATT GCAAACAGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC  
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC  
951 CCATTAGTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTAAACG  
10 1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAAACATAA AACTATTCT  
1051 CTACCTTATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT  
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA  
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGACAAA  
1201 GGAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGCGG  
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT  
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGATG GCAGTACCGT TACTTGGA  
1351 GTAAACGGCG TGGCAAACGA CCGCTGTCC AAAATCGGCA AAGGCACGCT  
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGACG  
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC  
15 1501 TTTAGTGAAA TCGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC  
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCCGGCTTT CCGCGCGGAC  
1601 GTTTGGATTT GAACGGGCAT TCGCTTCTGT TCCACCGCAT TCAAAATACC  
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT  
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCGGCAAT AACAACAAC  
20 1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT  
1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC  
1851 GGATCGCACT TTACTGCTTT CCGCGGAAC AAATTTAAAC GGCAATATCA  
1901 CGCAACAAAA CGGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC  
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCCAACAAGG  
25 2001 AGAAATCGTG TGGGACAACG ATTGGATCGA CCGCACATTT AAAGCGGAAA  
2051 ACTTCCATAT TCAGGCGGGA CAAGCGGTGG TTTCCCGCAA TGTGCGCAA  
2101 GTGGAAGGCG ATTGGCATT T AAGCAATCAC GCCCAAGCAG TTTTCGGTGT  
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC  
2201 TGACAAGTTG TACCGAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA  
30 2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCACGCTCA  
2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGGCG  
2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCAAAA CGGCAACCTC  
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAACCGG  
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACAACGCCG  
35 2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAGC  
2551 CATTCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA  
2601 TTTTGAAAAAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG  
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCCGG CACGGAATTA  
40 2701 GGCAATTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCTATCG  
2751 ACACGATGCG GCAGGCGCGC AAACCGGAG TGCGGAGAT GCGCCGCGCC  
2801 GCCGTTCCG CCGTTCCTTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA  
2851 TCCCGTTTCA ACACGCTGAC GGTAAACGGC AAATTGAACG GTCAGGAAAC  
2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC  
45 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC  
3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA  
3051 CACACCGCTG TCCGAAAATC TTAATTTCAC CCTGCaaaAc gaacacgtcg  
3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaagacgG CGAGTTCCgc  
3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc  
50 3201 gggagaaACA GAggcccT TGACGGCAAA ACAGGCacaA CTTGCCGCCA  
3251 AAcaacaggc ggaaaAAGAC AACgcgcaaa gccttgAcgc gctgattgcy  
3301 gCggggcgca atgccaccga AAAGGCagaa agtgttgccg aaccgGCCCCG  
3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA  
3401 AACGGGTGCA GGCGGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG  
55 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG  
3501 GGATTTGCCG CAACCGCAGC CCCAACCGCA ACCCAACCG CAGCGCGACC  
3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC  
3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA  
3651 CCGCCGCAAC GCCGTTTGGA CAAGCGGCAT CCGGGACACC AAACACTACC  
3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC  
60 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGCGATCC TGTTCGCA  
3801 CAACCGGACC GGAAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC

5  
10

```

3851 TTGCCACGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC
3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT
3951 CAGAGGCAAA ATCCGCCGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT
4001 ACCGCGCAGG TTTCGGCGGA TTCGGCATCG AACCACACAT CGGCGCAACG
4051 CGCTATTTTCG TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
4101 CACCCCGGGC CTTGCATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT
4151 ATTCAATCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTGAGCCTG
4201 TCCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGCG TCAATACCGC
4251 CGTATTGGCG CAGGATTTTCG GCAAAACCCG CAGTGCGGAA TGGGGCGTAA
4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCCAAG
4351 GGGCCGCAAT TGAAGCGCA GCACAGCGCG GGCATCAAAT TAGGCTACCG
4401 CTGGTAA

```

This is predicted to encode a protein having amino acid sequence (SEQ ID NO: 654):

15  
20  
25  
30  
35  
40  
45

```

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA RAGHTYFGIN
51 YQYYRDAEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNKYKAGT
151 NGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG
201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
251 KHSPYGLPT GGSPGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNFG
301 QLVRKDWFYD EIFAGDTHSV FYEPHQNGKY PFNDNNNGAG KIDAKHKHYS
351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDK
401 GKGEILITSN INQGAGGLYF EGNFTVSPKN NETWQGAGVH ISDGSTVTWK
451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQGGKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFGEKD
601 ATKTNGLNL NYPPEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLGSGWSK MEGIQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK
701 VEGDWHLSNH AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS
751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAETRTIRL RANATQNGNL
801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS
851 HSALNGNVSL ADKAVFHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGEL
901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRSL LSVTPPTSAAE
951 SRFNTLTVNG KLNQGQTFRF MSELFYRSG KLKLAESSEG TYTLAVNNTG
1001 NEPVSLQLT VVEGKDNTPL SENLNFTLQN EHVDAGAWRY QLIRKDGFR
1051 LHNVPKEQEL SDKLGKAGET EAALTAKQAAQ LAAKQQAED NAQSLDALIA
1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKKRVQADK DTALAKQREA
1151 ETRPATTAFP RARRARRDLP QPQPQPQP QRDLSRYAN SGLSEFSATL
1201 NSVFVQDEL DRVFAEDRRN AVWTSIGIRD KHYRSQDFRA YRQQTDLRQI
1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAFF GQYGIGRFDI
1301 GISAGAGFSS GSLSDGIRGK IRRRVLYGI QARYRAGFGG FGIEPHIGAT
1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL
1401 SYTDAASGKV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLHAAAAA
1451 GPQLEAQHSA GIKLGYRW*

```

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

ORF1-1 (SEQ ID NO: 650) and ORF1ng (SEQ ID NO: 654) show 93.7% identity in 1471 aa overlap:

50

```

          10      20      30      40      50      60
orf1-1.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN
          |||||

```

	orf1ng-1	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN
		10 20 30 40 50 60
5	orf1-1.pep	70 80 90 100 110 120 KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN      :
	orf1ng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN 70 80 90 100 110 120
10	orf1-1.pep	130 140 150 160 170 180 NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAPVEMTSYM      :
	orf1ng-1	NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSYM 130 140 150 160 170 180
15	orf1-1.pep	190 200 210 220 230 240 DGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG    :
	orf1ng-1	DGWKYADLNKYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG 190 200 210 220 230 240
20	orf1-1.pep	250 260 270 280 290 300 GTVNLGSEKIKHSPYGFLLPTGGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF      :
	orf1ng-1	GTVNLGSEKIKHSPYGFLLPTGGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF 250 260 270 280 290 300
25	orf1-1.pep	310 320 330 340 350 360 QLVRKDFWYDEIFAGDTHSVFYEPHQNGKYSFNDNNGTGKINAKHEHNSLPNRLKTRTV      :     :     :
	orf1ng-1	QLVRKDFWYDEIFAGDTHSVFYEPHQNGKYFFNDNNGAGKIDAKHKHYSLPYRLKTRTV 310 320 330 340 350 360
30	orf1-1.pep	370 380 390 400 410 420 QLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNIQAGAGGLYF      :
	orf1ng-1	QLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKKGELILTSNIQAGAGGLYF 370 380 390 400 410 420
35	orf1-1.pep	430 440 450 460 470 480 QGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGTLLHVQAKGENQGSIS :     :     :
	orf1ng-1	EGNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSVS 430 440 450 460 470 480
40	orf1-1.pep	490 500 510 520 530 540 VG DGT VILDQQADDKGKKQAFSEIGLVSGRGT VQLNADNQFNPDKLYFGFRGGRLDLNGH      :
	orf1ng-1	VGDGKVILDQQADDQGGKQAFSEIGLVSGRGT VQLNADNQFNPDKLYFGFRGGRLDLNGH 490 500 510 520 530 540
45	orf1-1.pep	550 560 570 580 590 600 SLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKD IATTGNNNSLDSKKEIAYNGWFGEKD      :
	orf1ng-1	SLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITTTGNNNNLDSKKEIAYNGWFGEKD 550 560 570 580 590 600
	orf1-1.pep	610 620 630 640 650 660 TTKTNGRLNLVYQPAEDRTL LLSGGT NLNGNITQTNGKLF FSGRPT PHAYNHLNDHWSQ

		610	620	630	640	650	660
	orf1ng-1	ATKTNGRLNLNYQPEEADRTLLSSGGTNLNGNITQTNGKLFSSGRPTPHAYNHLGSGWSK					
		670	680	690	700	710	720
5	orf1-1.pep	KEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNSNHAQAVFGVAPH					
	orf1ng-1	MEGIPQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNSNHAQAVFGVAPH					
		670	680	690	700	710	720
		730	740	750	760	770	780
10	orf1-1.pep	QSHTICTRSDWTGLTNCVEKTIITDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNL					
	orf1ng-1	QSHTICTRSDWTGLTSCTEKTIITDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLNGNL					
		730	740	750	760	770	780
		790	800	810	820	830	840
15	orf1-1.pep	SANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLT					
	orf1ng-1	SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLNNNAVQNGSLT					
		790	800	810	820	830	840
		850	860	870	880	890	900
20	orf1-1.pep	LSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTEL					
	orf1ng-1	LSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSGTEL					
		850	860	870	880	890	900
		910	920	930	940	950	960
25	orf1-1.pep	GNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFNTLT					
	orf1ng-1	GNLNLDNATITLNSAYRHDAAGAQTGSAADAPRRRSR---RSLLSVTPPTS AESRFNTLT					
		910	920	930	940	950	
		970	980	990	1000	1010	1020
30	orf1-1.pep	VNGKLNQGGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDN					
	orf1ng-1	VNGKLNQGGTFRFMSELFGYRSGKLKLAESSEGTYTLAVNNTGNEPVSLLEQLTVVEGKDN					
		960	970	980	990	1000	1010
		1030	1040	1050	1060	1070	
35	orf1-1.pep	KPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKA-----					
	orf1ng-1	TPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAGETEAALTAK					
		1020	1030	1040	1050	1060	1070
		1080	1090	1100	1110	1120	
40	orf1-1.pep	---EAKKQAEKDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAEEEKKRVO					
	orf1ng-1	QAQLAAKQAEKDNAQSLDALIAAGRNATEKAESVAEPARQAGGENAGIMQAEEEKKRVO					
		1080	1090	1100	1110	1120	1130
		1130	1140	1150	1160	1170	1180
45	orf1-1.pep	ADKDTALAKQREAE TRPATTAFPRARRARDLPQLQPQPQPQQRDLISRYANGLSEFS					
	orf1ng-1	ADKDTALAKQREAE TRPATTAFPRARRARDLPQPQPQPQPQQRDLISRYANGLSEFS					
		1140	1150	1160	1170	1180	1190
		1190	1200	1210	1220	1230	1240

	orf1-1.pep	ATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLG					
	orf1ng-1	ATLNSVFAVQDELDRVFAEDRRNAVVTSGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLG					
5		1200	1210	1220	1230	1240	1250
	orf1-1.pep	1250	1260	1270	1280	1290	1300
	orf1ng-1	1260	1270	1280	1290	1300	1310
10		SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGI					
	orf1-1.pep	1310	1320	1330	1340	1350	1360
	orf1ng-1	1320	1330	1340	1350	1360	1370
15		GKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR					
	orf1-1.pep	1370	1380	1390	1400	1410	1420
	orf1ng-1	1380	1390	1400	1410	1420	1430
20		AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI					
	orf1-1.pep	1430	1440	1450			
	orf1ng-1	1440	1450	1460			
25		KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX					
	orf1-1.pep	KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX					
	orf1ng-1	KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX					

25 In addition, ORF1ng (SEQ ID NO: 654) shows 55.7% identity with hap protein (P45387) (SEQ ID NO: 1153) over a 1455aa overlap:

SCORES Init1: 1104 Initn: 4632 Opt: 2680  
Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

30	orf1ng-1.pep	10	20	30	40	50	60
	p45387	MKTDDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYYRDFAE					
35		70	80	90	100	110	120
	orf1ng-1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN					
40		50	60	70	80	90	100
	p45387	KGKFTVGAQNIKVYNKQQLVGTSMTKAPMIDFSVSRNGVAALVENQYIVSVAHNVGYT					
45		130	140	150	160	170	180
	orf1ng-1.pep	NVDFGAEGSNPDQHRFSYQIVKRNKYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSYM					
50		110	120	130	140	150	160
	p45387	DVDFGAEGNPDQHRFTYKIVKRNKYKD-NLHPYEDDYHNPRLHKFVTEAAPIDMTSMN					
55		190	200	210	220	230	240
	orf1ng-1.pep	DGWKYADLNKYPDRVRIGAGRQYWRSDDEPNRESSYHIASAYSWLVGGNTFAQNGSGG					
60		170	180	190	200	210	220
	p45387	NGSTYSDRTKYPERVRIGSGRQFWRNDQDKG-----QVAGAYHYLTAGNTHNQRGAGN					

		170	180	190	200	210		
		250	260	270	280	290	300	
5	orflng-1.pep	GTVNLGSEKIKHSPYGF	LPTGGSF	GDGSPMFI	YDAQKQKWL	INGVLQTGN	PNYIGKSN	
	p45387	GYSYLG	GGDVRKAGEY	GPLPIAGS	KGDSGSPMFI	YDAEKQKWL	INGILREGNPF	
		220	230	240	250	260	270	
		310	320	330	340	350	360	
10	orflng-1.pep	QLVRKDW	FYDEIFAGD	THSVFYEP	HQNGKYFF	NDNNNGAG	KIDAKHKH	
	p45387	QLVRKSYF	-DEIFERDL	HTSLYTRAG	NGVYTISG	NDNGQGS	ITQKS---	
		280	290	300	310	320		
		370	380	390	400	410	419	
15	orflng-1.pep	QLFNVS	LSETAREP	VYHAA-GG	VNSYRPR	LNNNGEN	ISFIDK	
	p45387	TLANMS	LPLKEKDK	VHNPRYD	GPNIYSP	RLNNGET	LYFMDQ	
		330	340	350	360	370	380	
		420	430	440	450	460	470	479
20	orflng-1.pep	FEGNFT	VSPKNNET	WQAGVHIS	DGSTVTW	KVNGVAN	DRLSKIG	
	p45387	FEGNFT	VSPNSNQ	TWQAGAI	HVSENST	VTWVNG	VEHDRLS	
		390	400	410	420	430	440	
		480	490	500	510	520	530	539
25	orflng-1.pep	SVGDGK	VILDQQAD	DQGGKQAF	SEIGLV	SGRGT	VQLNAD	
	p45387	SVGDGK	VILEQQAD	DQGNKQAF	SEIGLV	SGRGT	VQLNDK	
		450	460	470	480	490	500	
		540	550	560	570	580	590	
30	orflng-1.pep	HSLSF	HRIQNTDE	GAMIVNHN	QDKESTV	ITGNKD	ITT-TG	
	p45387	HSLTF	KRIQNTDE	GAMIVNHN	TQAANVT	ITGNES	IVLPNG	
		510	520	530	540	550	560	
		600	610	620	630	640	650	
35	orflng-1.pep	EKDATK	TNGRLNL	NYQPEE	ADRTLL	SSGGTN	LN	
	p45387	ETDKN	KHNGRL	NLIYKP	TTEDRT	LLSSGG	TN	
		570	580	590	600	610	620	
		660	670	680	690	700	710	
40	orflng-1.pep	WSKME	GIPQGEI	VWDNDW	IDRTF	KAENFH	IQQGQ	
	p45387	WSEME	GIPQGEI	VWDHDW	INRTF	KAENF	QIKGGS	
		630	640	650	660	670	680	
		720	730	740	750	760	770	
45	orflng-1.pep	APHQS	HTICTR	SDWTGL	TSCTEK	ITDDK	VIASL	
	p45387	VPNQ	QNTICT	RSWTGL	TTCQK	VDLTD	TKVINS	
		690	700	710	720	730	740	
		780	790	800	810	820	830	
	orflng-1.pep	GNLSAG	GDTHY	TVTRN	ATQNGN	LSLVG	NAQAT	



		p45387	GNVTL-----TNHSQFTLSNNATQIG
		750	760 770
5	orf1ng-1.pep	840 850 860 870 880 890	SLTLSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDESWTLPSG
	p45387	780 790 800 810 820 830	NIRLSDNSTATVDNANLNGNVHLTDSAQFSLKNSHFSHQIQGDKGTTVTLENATWTMPSD
10	orf1ng-1.pep	900 910 920 930 940 950	TELGNLNLNDNATITLNSAYRHDAAGAQTGSAADAPRRRSRRSLLSVTPPTSASERFNTLT
	p45387	840 850 860 870	TTLQNLTLLNSTITLNSAY-----SASSNNTPRRRS---LETETTPTSAEHRFNTLT
15	orf1ng-1.pep	960 970 980 990 1000 1010	VNGKLNQGTFRFRMSELFGRSGYRSGKLKLAESSEGTYYTLAVNNTGNPVSLEQLTVVEGKDN
	p45387	880 890 900 910 920 930	VNGKLSGQGTFOFTSSLFGYKSDKLKLSNDAEGDYILSVRNTGKEPETLEQLTLVESKDN
20	orf1ng-1.pep	1020 1030 1040 1050 1060 1070	TPLSENLFNFTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLGKAGETEAAALTAK
	p45387	940 950 960 970 980 990	QPLSDKLFKFTLENDHVDAGALRYKLVKNDGFEFLHNPVKEQELHNDLVRAEQAERTLEAK
25	orf1ng-1.pep	1080 1090 1100 1110 1120 1130	QAQLAAKQQAQEKDQSLDALIAAGRNAT-EKAESVAEPARQAGGENAGIMQAEEEKRV
	p45387	1000 1010 1020 1030 1040 1050	QVEPTAKTQTGEFKVRSRRAAFAFPDTPDQSLNLALEAKQAE-LTAETQKSKAKTKKV
30	orf1ng-1.pep	1140 1150 1160 1170 1180 1190	QADK---DTALAKQREAEATRPATTAFPRARRARRD-LPQPQPQPQPQORDLISRYANS
	p45387	1060 1070 1080 1090 1100 1110	RSKRAVFSDDLQSLFALEAALEVIDAPQQSEKDRLAQEEAEKQ-RKQKDLISRYNSA
35	orf1ng-1.pep	1200 1210 1220 1230 1240 1250	LSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQQ-TDLRQIG
	p45387	1120 1130 1140 1150 1160 1170	LSELSATVNSMLSVQDELDRFVDQAQSAVWTNIAQDKRRYDSDAFRAYQQKTNLRQIG
40	orf1ng-1.pep	1260 1270 1280 1290 1300 1310	MQKNLGSGRVGILFVSHNRTGNFTDDGIGNSARLAHGAVFGQYIGRFDIGISAGAGFSSG
	p45387	1180 1190 1200 1210 1220 1230	VQKALANGRIGAVFVSHSRSDNTFDEQVKNHATLTMMSGFAQYQWGDQFGVNVGTGISAS
45	orf1ng-1.pep	1320 1330 1340 1350 1360 1370	SLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGL
	p45387	1240 1250 1260 1270 1280 1290	KMAEEQSRKIHRKAINYGVNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEVVRKTPSL
		1380 1390 1400 1410 1420 1430	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEW

```

      ||||| |||::||:| |::||: ||: ::||:::|:| || :|| | ||: :|
p45387 AFNRYNAGIRVDYTFPTDNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEV
      1300      1310      1320      1330      1340      1350

      1440      1450      1460      1469
5      orflng-1.pep  GVNAEIKGFTLSLHAAAAKGPQLEAQHSAGIKLGYRWX
      ||:|||| | :| : ::| || |::|:|||||
p45387 GLKAEILHFQISAFISKSQGSQGLKQONVGVKLGYSRW
      1360      1370      1380      1390

```

- 10 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 78

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 655):

```

      1  ..AAGGTGTGGC AATTTGTCGA AGA.CCGCTG CGTGCCGTCG TGCCTGCCGA
15      51  CAGTTTGTAA CCGACCGCGC AAAAAATTGAA CCTGTTTAAG GCGGGTGCGG
      101  CAACCATTTT GTTTTATGAA GATCAAAATG TCGTCAAAGG TTTGCAGGAG
      151  CAGTTCCTTG CTTATGCCGC TAACCTCCCC GTTTGGGCGG ATCAGGCAAA
      201  CGCGATGGTG CAGTATGCCG TTTGGACGAC ACTTGCCGCG GTCGGCGTAG
20      251  GTGCAACCT GCAACATTAC AATCCCTTGC CCGATGCGGC GATTGCCAAA
      301  GCGTGGAATA TCCCCGAAAA CTGGTTGTTG CGCGCACAAA TGGTTATCGG
      351  CCGTATTGAA GGGGCGGCAG GTGAAAAGAC CTTTGAACCC GTTGCAGAAC
      401  GTTTGAAAGT GTTCGGCGCA TAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 656; ORF6):

```

      1  ..KVVQFVEXPL RAVVPADSFE PTAQKLNLFK AGAATILFYE DQNVVKGLQE
25      51  QFPAYAAANFP VWADQANAMV QYAVWTTLAA VGVGANLQHY NPLPDAAIAK
      101  AWNIPENWLL RAQMVIIGIE GAAGEKTFEP VAERLKVFGA *

```

Further sequence analysis revealed a further partial DNA sequence (SEQ ID NO: 657):

```

      1  ..CTGCGTGCCG TCGTGCCTGC CGACAGTTTT GAACCGACCG CGCAAAAATT
30      51  GAACCTGTTT AAGGCGGGTG CGGCAACCAT TTTGTTTAT GAAGATCAAA
      101  ATGTCGTCAA AGGTTTGAG GAGCAGTTCC CTGCTTATGC CGCTAACTTC
      151  CCCGTTTGGG CGGATCAGGC AAACGCGATG GTGCAGTATG CCGTTTGGAC
      201  GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT TACAATCCCT
      251  TGCCCGATGC GCGGATTGCC AAAGCGTGGA ATATCCCGA AACTGGTTG
35      301  TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG CAGGTGAAAA
      351  GACCTTTGAA CCCGTTGCAG AACGTTTGAA AGTGTTCCGGC GCATAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 658; ORF6-1):

```

      1  ..LRAVVPADSF EPTAQKLNLF KAGAATILFY EDQNVVKGLQ EQFPAYAAANF
40      51  PVWADQANAM VQYAVWTTLA AVGVGANLQH YNPLPDAAIA KAWNIPENWL
      101  LRAQMVIGGI EGAAGEKTFE PVAERLKVFG A*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

[illegible]

20	1	ATGACCCGTC	AATCTCTGCA	ACAGGCTGCC	GAAAGCCGCC	GTTCCATTTA
	51	TTCGTAAAT	AAAAATCTGC	CCGTCGGCAA	AGATGAAATC	GTCCAAATCG
	101	TCGAACACGC	CGTTTTGCAC	ACACCTTCTT	CGTTCAATTC	CCAATCTGCC
	151	CGTGTGGTCG	TGCTGTTTGG	CGAAGAGCAT	GATAAGGTGT	GGCAATTTGT
25	201	CGAAGACGCG	CTGCGTGCCG	TCGTGCCTCG	CGACAGTTTT	GAACCCGACG
	251	CGCAAAAATT	GAACCTGTTT	AAGCGGGGTG	CGGCCAACTAT	TTTGTTTTTAT
	301	GAAGATCAAA	ATGTCGTCAA	AGGTTTGAG	GAGCAGTTCC	CTGCTTATGC
	351	CGCCAACCTT	CCCGTTTGGG	CGGACCAGGC	GAACGCGATG	GTGCAGTATG
30	401	CCGTTTGGAC	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT
	451	TACAATCCCT	TGCCCAGATG	GGCGATTGCC	AAAGCGTGGA	ATATCCCCGA
	501	AAACTGGTTG	TTGCGCGCAC	AAATGTTTAT	CGGCGGTATT	GAAGGGGGCG
	551	CAGGTGAAAA	GACCTTTGAA	CCAGTTGCAG	AACGTTTGAA	AGTGTTCGGC
	601	GCATAA				

35

1	MTRQSLQQA	ESRRSYSLN	KNLPVGKDEI	VQIVEHAVLH	TPSSFNSQSA
51	RVVVLFGEEH	DKVWQFVEDA	LRAVVPADSF	EPTAQKLNLF	KAGAATILFY
101	EDQNVVKGLQ	EQFPAYANF	PVWADQANAM	VQYAVWTTLA	AVGVGANLQH
151	YNPLPDAAIA	KAWNIPENWL	LRAQMVIIGI	EGAAGEKTFE	PVAERLKVFG
201	A*				

40 ORF6a (SEQ ID NO: 660) and ORF6-1 (SEQ ID NO: 658) show 100.0% identity in 131 aa overlap:

45

	50	60	70	80	90	100
orf6a.pep	TPSSFNSQSARVVVLFGE	EHDKVWQFVEDALRA	VVPADSF	EP	TAQKLNLF	KAGAATILFY
orf6-1						

		110	120	130	140	150	160
	orf6a.pep	EDQNVVKGLQE	QFPAYAA	NFPVWADQ	ANAMVQYAV	WTTTAAV	GVGANLQHY
	orf6-1	EDQNVVKGLQE	QFPAYAA	NFPVWADQ	ANAMVQYAV	WTTTAAV	GVGANLQHY
5		40	50	60	70	80	90
		170	180	190	200		
	orf6a.pep	KAWNIPENW	LLRAQMVI	GIEGAAGE	KTFEPVAE	RLKVFGAX	
	orf6-1	KAWNIPENW	LLRAQMVI	GIEGAAGE	KTFEPVAE	RLKVFGAX	
10		100	110	120	130		

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 (SEQ ID NO: 656) shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) (SEQ ID NO: 662) from *N.gonorrhoeae*:

	orf6.pep	KVWQFVEXPLRAVVPADSFEPTAQKLNLFK	30
15	orf6ng	SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLFK	64
	orf6.pep	AGAATILFYEDQNVVKGLQE	90
	orf6ng	AGAATILFYEDQNVVKGLQE	124
20	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVI	140
	orf6ng	NPLPDVAIAKAWNIPENWLLRAQMVI	174

The complete length ORF6ng nucleotide sequence (SEQ ID NO: 661) was identified as:

	1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGT
25	51	ACGCATGGGA	TTACCCTTAT	ATATTGCGTC	CCTAAGAAGG	GGCGCAATAT
	101	ATAAGGTGTG	GCAATTGTG	GAAGACGCGC	TGCGTGCCGT	CGTGCCTGCC
	151	GACAGTTTGT	AACCGACCGC	GCAAAATTG	AAGCTGTTTA	AGGCGGGCGC
	201	GGCAACCATT	TTGTTTATG	AAGATCAAAA	TGTCGTCAA	GGTTGCAGG
	251	AGCAGTCCC	TGCTTATGCC	GCCAACTTTC	CCGTTGGGC	GGACCAAGCG
30	301	AACGCTATGG	TACAGTATGC	CGTCTGGACG	ACACTTGCCG	CGGTCGGTGC
	351	AGGTGCAAAT	CTGCAACATT	ACAACCCCTT	GCCCGATGTG	GCGATTGCTA
	401	AAGCGTGGAA	TATCCCGAA	AACTGGCTGT	TGCGCGCGCA	AATGGTTATC
	451	GGTGGTATTG	AAGGGGcggc	aggtgaaaaa	gtctttgaac	CCGTTGCgga
35	501	acgtttgAAA	GTGTTCCGCG	CATAA		

This encodes a protein having amino acid sequence (SEQ ID NO: 662):

	1	MAVASNVSLD	MSNPTVLRMG	LPLYIASLRR	GAIYKVWQFV	EDALRAVVPA
	51	DSFEPTAQKL	KLFKAGAATI	LFYEDQNVVK	GLQE	QFPAYANFPVWADQA
	101	NAMVQYAVWT	TLAAVGAGAN	LQHYNPLPDV	AIKAWNIP	NWLLRAQMVI
40	151	GGIEGAAGEK	VFEPVAERLK	VFGA*		

ORF6ng (SEQ ID NO: 662) and ORF6-1 (SEQ ID NO: 658) show 96.9% identity in 131 aa overlap:

orf6-1.pep  
 LRAVVPADSFEP<sup>10</sup>TAQKLNLFKAGAATILFY  
 orf6ng  
 PTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEP<sup>20</sup>TAQK<sup>30</sup>KL<sup>40</sup>LFKAGAATILFY  
 orf6-1.pep  
 EDQNVVKGLQE<sup>40</sup>QFPAYAANFPVWADQANAMVQYAVWTTLA<sup>50</sup>AVGVGANLQHYNPLPDAAIA  
 orf6ng  
 EDQNVVKGLQE<sup>80</sup>QFPAYAANFPVWADQANAMVQYAVWTTLA<sup>90</sup>AVGAGANLQHYNPLPDVAIA  
 orf6-1.pep  
 KAWNIPENWLLRAQM<sup>100</sup>VIGGIEGAAGEKTFEPVAERLKVFGAX  
 orf6ng  
 KAWNIPENWLLRAQM<sup>140</sup>VIGGIEGAAGEKVFE<sup>150</sup>PVAERLKVFGAX

It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 79

The following partial DNA sequence was identified in *N.meningitidis* (SEQ.ID NO: 663)

20	1	..GGCTACAACT	ACCTGTTCGC	GCGCGGCAGC	CGCATCGCCA	ACTACCAAAT
	51	CAACGGCATC	CCCGTTGCCG	ACGCGCTGGC	CGATACGGG <u>t</u>	CAATGCCAAC
	101	ACCGCCGCCT	ATGAGCGCGT	AGAAGTCGTG	CGCGCGGTGG	CGGGGCTGCT
	151	GGACGGCAGC	GGCGAGCCTT	CCGCCACCGT	CAATCTGGTG	CGCAAACGCG
25	201	TGACCCGCAA	GCCATTGTTT	GAAGTCCGCG	CCGAAGCgGG	CAACCGcAAA
	251	CATTTCGGGC	TGGACGCGGA	CGTATCCGGC	AGCCTGAACA	CCGAAg.crc
	301	rCTGCGCgGC	CGCCTGGTTT	CCAcCTTCGG	ACGCGCGCAG	TCGTGGCGGC
	351	GGCGCGAACG	CAGCCGskAT	GCCGAActCT	ACGGCATTTT	GGAATACGAC
30	401	ATCGCACCGC	AAACCCGCGT	CCACGCArgC	ATGGACTACC	AGCAGGCGAA
	451	AGAAACCGCC	GACGCGCCGC	TCAGcTACGC	CGTGTAGCAG	AGCCAAGGTT
	501	ATGCCACCGC	CTTCGGCCCC	AAAGACAACC	CCGCCACAAA	TTGGGCGAAC
	551	AGCCACACC	GTGCGCTCAA	CCTGTTCGCC	GGCATCGAAC	ACCGCTTCAA
	601	CCAAGACTGG	AAACTCAAAG	CCGAATACGA	CTAC..	

This corresponds to the amino acid sequence (SEQ ID NO: 664; ORF23):

```

35      1      . . GYN Y L F A R G S   R I A N Y Q I N G I   P V A D A L A D T G   N A N T A A Y E R V   E V V R G V A G L L
      51      D G T G E P S A T V   N L V R K R L T R K   P L F E V R A E A G   N R K H F G L D A D   V S G S L N T E X X
     101      L R G R L V S T F G   R G D S W R R R E R   S R X A E L Y G I L   E Y D I A P Q T R V   H A X M D Y Q Q A K
     151      E T A D A P L S Y A   V Y D S Q G Y A T A   F G P K D N P A T N   W A N S H H R A L N   L F A G I E H R F N
     201      Q D W K L K A E Y D   Y . . .
40

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 665):

45

1	ATGACACGCT	TCAATATTC	CCTGCTGTTT	GCCGCCCTGT	TGCCCGTGA
51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
101	CTGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGACCCGC	GAGTTCCAAC
151	GACGGCTACA	CTGTTTCCGG	CACGCACACC	CCGCTCGGGC	TGCCCATGAC

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40

```

201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GGCGACCGGC
301 ACCAGCCGCC AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT
351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG
401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
451 GTAGAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT
551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGACGCG
601 GACGTATCGG GCAGCTGAA CACCGAAGGC ACGCTGCGCG GCCGCTGGT
651 TTCCACCTTC GGACGCGCGG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG
701 ATGCCGAACCT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC
751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCGCTC
901 AACCTGTTCG CCGGCATCGA ACACCGCTTC AACCAGACT GGAAACTCAA
951 AGCCGAATAC GACTACACC GCAGCCGCTT CCGCCAGCCC TACGGCTAG
1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
1051 GGTATTATGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCAATTGAT
1101 CGGCAAATAC CGCCTGTTTCG GCCGCGAACA CGATTTAATC GCGGGTATCA
1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC
1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA
1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG
1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG
1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC
1501 AGCCTGTTTCG TCCCAGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG
1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCCGGCCG TAAAAACAAC
1651 CTGCCACCG CAGCAGGAGC CGACCCGAGC GGCAACACCT ACTACCGCGC
1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
1801 GACCAAGACG GCAGCCGCCT GAACCCGAC AGCGTACCCG AACCGAGCTT
1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC
1951 ACGTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCG CCGACAACAG
2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCGC CGAACTGTCG CTGAACGTGG ACAATCTGTT CAACAAACAC
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
2151 CGCGGCGTTT ACCTATCGGT TTAAATAA
  
```

This corresponds to the amino acid sequence (SEQ ID NO: 666; ORF23-1):

45  
 50  
 55

```

1  MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN
51  DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
101 TSRQIYGSDR AGYNYLFARG SRIANYQING IPVADALADT GNANTAAYER
151 VEVVRGVAGL LDGTGEPSTF VNLVRKRLTR KPLFEVRAEA GNRKHFGLDA
201 DVSGSLNTEG TLRGRLVSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQOP YGVAGVLSID HNTAATDLIP
351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
451 ILGGRYTRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
501 SLFVPQSQKD EHGSYLPVPT GNNLEAGIKG EWLEGRNLNAS AAVYRARKNN
551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHDPA
651 TLRIPNPAAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVDNLPNKH
701 YRTQPDHRSY GALRTVNAAF TYRFK*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047) (SEQ ID NO: 1154)

ORF23 (SEQ ID NO: 664) and PupB protein (SEQ ID NO: 1154) show 32% aa identity in 205aa overlap:

```

5      Orf23  6  FARGSRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRK 65
      ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
      PupB   215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273

      Orf23  66  RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXXXXXXAE 125
      R T +      + EAGN +G DVSG L +RGR V+ +
10      PupB   274 RPTAEAQASITGEAGNWDRYGTGFDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333

      Orf23  126 LYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
      +YGI E+D++ T + Y + D+PL + S G T N A +W+
      PupB   334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSLNAAPDWSY 391

      Orf23  184 SHHRALNLFAGIEHRFNQDWKLKAE 208
      + H + F IE + W K E
15      PupB   392 NDHEQTSFFTSIEQQLGNGWSGKIE 416
  
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 (SEQ ID NO: 664) shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) (SEQ ID NO: 668) from strain A of *N. meningitidis*:

```

20      orf23.pep                                10      20      30
      GYNYLFARGSRIANYQINGIPVADALADTG
      orf23a  QMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARGSRIANYQINGIPVADALADTG
      90      100      110      120      130      140

25      orf23.pep      40      50      60      70      80      90
      NANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD
      orf23a  NANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTRKPLFEVRAEAGNRKHFGLGAD
      150      160      170      180      190      200

30      orf23.pep      100      110      120      130      140      150
      VSGSLNTEXXLRGRLVSTFGRGDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQA
      orf23a  VSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQA
      210      220      230      240      250      260

35      orf23.pep      160      170      180      190      200      210
      ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD
      orf23a  ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD
      270      280      290      300      310      320
  
```

orf23.pep      Y  
 |  
 orf23a      YTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIA  
                                  330                   340                   350                   360                   370                   380

The complete length ORF23a nucleotide sequence (SEQ ID NO: 667) is:

```

1  ATGACACGCT  TCAAATATTC  CCTGCTGTTT  GCCGCCCTGT  TGCCCGTGTA
51  CGCGCAGGCC  GATGTTTCTG  TTTCAGACGA  CCCAAAACCG  CAGGAAAGCA
101 CTGAATTGCC  GACCATCACC  GTTACCGCCG  ACCGCACCGC  GAGTTCCAAC
151 GACGGCTACA  CTGTTTCCGG  CACGCACACC  CCGCTCGGGC  TGCCCATGAC
201 CCTGCGCGAA  ATCCCGCAGA  GCGTCAGCGT  CATCACATCG  CAACAAATGC
251 GCGACCAAAA  CATCAAAGCG  CTCGACCGCG  CCCTGTTGCA  GGCGACCGGC
301 ACCAGCCGCC  AGATTTACGG  CTCCGACCGC  GCGGGCTACA  ACTACCTGTT
351 CGCGCGCGGC  AGCCGCATCG  CCAACTACCA  AATCAACGGC  ATCCCGTTG
401 CCGACGCGCT  GGCGATACG  GGCAATGCCA  ACACCGCCGC  CTATGACGCG
451 GTAGAAGTCG  TCGCGGGCGT  GGCGGGGCTG  CTGGACGGCA  CGGGCGAGCC
501 TTCCGCCACC  GTCAATCTGG  TGCGCAAACG  CCCGACCCGC  AAGCCATGT
551 TTGAAGTCCG  CGCCGAAGCG  GGCAACCGCA  AACATTTCCG  GCTGGGCGCG
601 GACGTATCGG  GCAGCCTGAA  TGCCGAAGGC  ACGCTGCGCG  GCCGCTGGT
651 TTCCACCTTC  GGACGCGGCG  ACTCGTGGCG  GCAGCGCGAA  CGCAGCCGCG
701 ATGCCGAACT  CTACGGCATT  TTGGAATACG  ACATCGCACC  GCAAACCCGC
751 GTCCACGCAG  GCATGGACTA  CCAGCAGGCG  AAAGAAAACCG  CCGACGCGCC
801 GCTCAGCTAC  GCCGTGTACG  ACAGCCAAGG  TTATGCCACC  GCCTTCGGCC
851 CGAAAGACAA  CCCC GCCACA  AATTGGGCGA  ACAGCCGCCA  CCGTGCGCTC
901 AACCTGTTCG  CCGGCATCGA  ACACCGCTTC  AACCAAGACT  GGAAACTCAA
951 AGCCGAATAC  GACTACACCC  GCAGCCGCTT  CCGCCAGCCC  TACGGCGTAG
1001 CAGGCGTGCT  TTCCATCGAC  CACAACACCG  CCGCCACCGA  CCTGATTCCC
1051 GGTATTATGG  ACGCCGACCC  GCGCACCCAC  AGCGCCAGCG  TGTCATTAAT
1101 CGGCAAAATAC  CGCCTGTTTC  GCCGCGAACA  CGATTTAATC  GCGGGTATCA
1151 ACGGTTACAA  ATACGCCAGC  AACAATACG  GCGAACGCAG  CATCATCCCC
1201 AACGCCATTC  CCAACGCCTA  CGAATTTTCC  CGCACGGGTG  CCTACCCGCA
1251 GCCTGCATCG  TTTGCCCAAA  CCATCCCGCA  ATACGGCACC  AGGCGGCAAA
1301 TCGGCGGCTA  TCTCGCCACC  CGTTTCCGCG  CCGCCGACAA  CCTTCGCTG
1351 ATACTCGGCG  GCAGATACAG  CCGTTACCGC  ACCGGCAGCT  ACGACAGCCG
1401 CACACAAGGC  ATGACCTATG  TGTCCGCCAA  CCGTTTCACC  CCCTACACAG
1451 GCATCGTGTT  CGACCTGACC  GGCAACCTGT  CGTTTACGG  CTCGTACAGC
1501 AGCCTGTTTC  TCCCGCAATC  GCAAAAAGAC  GAACACGGCA  GCTACCTGAA
1551 ACCCGTAACC  GGCAACAATC  TGGAAGCCGG  CATCAAAGGC  GAATGGCTTG
1601 AAGGCCGTCT  GAACGCATCC  GCCGCCGTGT  ACCGCGCCCG  TAAAAACAAC
1651 CTCGCCACCG  CAGCAGGACG  CGACCCGAGC  GGCAACACCT  ACTACCGCGC
1701 CGCCAACCAA  GCCAAAACCC  ACGGCTGGGA  AATCGAAGTC  GGCGGCCGCA
1751 TCACGCCCGA  ATGGCAGATA  CAGGCAGGTT  ACAGCCAAAG  CAAAACCCGC
1801 GACCAAGACG  GCAGCCGCCT  GAACCCCGAC  AGCGTACCCG  AACGCAGCTT
1851 CAAACTCTTC  ACTGCCTACC  ACTTTGCCCC  CGAAGCCCCC  AGCGGCTGGA
1901 CCATCGGCGC  AGGCGTGCGC  TGGCAGAGCG  AAACCCACAC  CGACCCTGCC
1951 ACGCTCCGCA  TCCCCAACCC  CGCCGCCAAA  GCCGCGCCCG  CCGACAACAG
2001 CCGCCAAAAA  GCCTACGCCG  TCGCCGACAT  CATGGCGCGT  TACCGCTTCA
2051 ATCCGCGCGC  CGAAGTGTG  CTGAACGTGG  ACAATCTGTT  CAACAAACAC
2101 TACCGCACCC  AGCCCGACCG  CCACAGCTAC  GGCGCACTGC  GGACAGTGAA
2151 CGCGGCGTTT  ACCTATCGGT  TTAAATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 668):

```

1  MTRFKYSLLF  AALLPVYAQA  DVSVSDDPKP  QESTELPTIT  VTADRTASSN
55  DGYTVSGTHT  PLGLPMTLRE  IPQSVSVITS  QQMRDQNIKA  LDRALLQATG
101  TSRQIYGS DR  AGYNLFFARG  SRIANYQING  IPVADALADT  GNANTAYER
151  VEVVRGVAGL  LDGTGEPSAT  VNLVRKRPR  KPLFEVRAEA  GNRKHFGLGA
201  DVSGSLNAEG  TLRGRLVSTF  GRGDSWRQRE  RSRDAELYGI  LEYDIAPQTR

```



5  
10  
15  
20  
25  
30  
35  
40  
45  
50

```

251  VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301  NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP
351  GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401  NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
451  ILGGYRSRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
501  SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRLNAS AAVYRARKNN
551  LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601  DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHTDPA
651  TLRIPNPAAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH
701  YRTQPDRHSY GALRTVNAAF TYRFK*

```

ORF23a (SEQ ID NO: 668) and ORF23-1 (SEQ ID NO: 666) show 99.2% identity in 725 aa overlap:

15  
20  
25  
30  
35  
40  
45  
50

```

      10      20      30      40      50      60
orf23a.pep  MTRFKYSLLF AALLPVYAQADSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
orf23-1      MTRFKYSLLF AALLPVYAQADSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
      10      20      30      40      50      60

      70      80      90     100     110     120
orf23a.pep  PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNLYFARG
orf23-1      PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNLYFARG
      70      80      90     100     110     120

      130     140     150     160     170     180
orf23a.pep  SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTTR
orf23-1      SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTR
      130     140     150     160     170     180

      190     200     210     220     230     240
orf23a.pep  KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGRGDSWRQRERSRDAELYGI
orf23-1      KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGLVSTFGRGDSWRRRERSRDAELYGI
      190     200     210     220     230     240

      250     260     270     280     290     300
orf23a.pep  LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL
orf23-1      LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL
      250     260     270     280     290     300

      310     320     330     340     350     360
orf23a.pep  NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
orf23-1      NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
      310     320     330     340     350     360

      370     380     390     400     410     420
orf23a.pep  SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS
orf23-1      SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS
      370     380     390     400     410     420

      430     440     450     460     470     480
orf23a.pep  FAQTIPQYGT RRQIGGYLAT RFRAADNLSLILGGYRSRYRTGSYDSRTQGMTYVSANRFT

```

-484-

	orf23-1	 FAQTIPQYGTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMYVSANRFT	430	440	450	460	470	480
5	orf23a.pep	490 500 510 520 530 540 PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS						
	orf23-1	 PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS	490	500	510	520	530	540
10	orf23a.pep	550 560 570 580 590 600 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR						
	orf23-1	 AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR	550	560	570	580	590	600
15	orf23a.pep	610 620 630 640 650 660 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK						
	orf23-1	 DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK	610	620	630	640	650	660
20	orf23a.pep	670 680 690 700 710 720 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF						
	orf23-1	 ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF	670	680	690	700	710	720
25	orf23a.pep	TYRFXK						
	orf23-1	 TYRFXK						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 (SEQ ID NO: 664) shows 93.4% identity over a 211aa overlap with a predicted ORF  
 30 (ORF23.ng) (SEQ ID NO: 670) from *N. gonorrhoeae*:

	orf23.pep	GYNLYFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNLYFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
35	orf23.pep	GTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFG LDADVSGSLNTEXXLRGRLVSTFGR	111
	orf23ng	GTGEP SATVNLVRKHPTRKPLFEVRAEAGNRKHFG LGADVSGSLNAEGLTRGRLVSTFGR	120
	orf23.pep	GDSWRRRERSR XAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSR DAELYGILEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAF	180
40	orf23.pep	GPKDNPATN WANS HHRLNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATN WSNRNRALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence (SEQ ID NO: 669) is predicted to encode a protein comprising amino acid sequence (SEQ ID NO: 670):

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEP SATVN	LVRKHPTRKP	LFEVRAEAGN	RKHFGLGADV
5	101	SGSLNAEGL	RGR LVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRNRALNL
	201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPQPSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
10	351	GGRYSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
	401	FVPQLQKDEH	GSYLKPV TGN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQA	GYSQSKPRDQ
	501	DGSRLNPDSV	PERSFKLFTA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLFNKHRY
15	601	TQPDRHSYGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence (SEQ ID NO: 671):

	1	ATGACACGCT	TCAAATACTC	CCTGCTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
20	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCGAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
25	351	CGCGCGCGGC	AGCCGCATCG	CCAACTACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTTCG	GCTGGGCGCG
30	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
	751	GTCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
35	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCcgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
40	1101	CGGCAAATAC	CgcctGTTTCG	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTTCC
	1201	AACGCCATT	CCAACGCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
45	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG
	1401	CACACAAGGC	ATGACCTATG	TGTCCGCCAA	CCGTTTCACC	CCCTACACAG
	1451	GCATCGTGTT	CGATCTGACC	GGCAACCTGT	CGCTTTACGG	CTCGTACAGC
	1501	AGCCTGTTCG	TCCC GCAATT	GCAAAAAGAC	GAACACGGCA	GCTACCTGAA
	1551	ACCCGTAACC	GGCAACAATC	TGGAAGCCGA	CATCAAAGGC	GAATGGCTTG
50	1601	AAGGGCGTCT	GAACGCATCC	GCCGCCGTGT	ACCGCGCCCG	TAAAAACAAC
	1651	CTCGCCACCG	CAGCAGGACG	CGACCAGAGC	GGCAACACCT	ACTATCGCGC
	1701	CGCCAACCAA	GCCAAAACCC	ACGGCTGGGA	AATCGAAGTC	GGCGGCCGCA
	1751	TCACGCCCGA	ATGGCAGATA	CAGGCAGGCT	ACAGCCAAG	CAAAACCCGC
	1801	GACCAAGACG	GCAGCCGCCT	GAACCCCGAC	AGCGTAcCCG	AACGCAGCTT
55	1851	CAAACCTCTT	ACCGCCTACC	ACTTAGCCCC	CGAAGCCCCC	AGCGGCCGGA
	1901	CCATcggtGC	GGGTGTGCGC	CGGCAGGGCG	AAACCCACAC	CGACCCAGCC

5  
1951 GCGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG  
2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
2051 ATCCGCGCAC CGAACTGTCG CTGAACGTGG ACAACCTGTT CAACAAACAC  
2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCGACTGC GGACAGTGAA  
2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence (SEQ ID NO: 672; ORF23ng-1):

10  
15  
20  
1 MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN  
51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
101 TSRQIYGSDR AGYNYLFARG SRIANYQING IPVADALADT GNANTAAYER  
151 VEVVRGVAGL PDGTGEP SAT VNLVRKHPTR KPLFEVRAEA GNRKHFGLGA  
201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR  
251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSNRNAL  
301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQRP YGVAGVLSID HSTAATDLIP  
351 GYWHADPRTH SASMSLTGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
401 NAIPNAYEFS RTGAYPQPSS FAQTIPQYDT RRQIGGYLAT RFRAADNLSL  
451 ILGGYRSYR AGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS  
501 SLFVPQLQKD EHGSYLKPVT GNNLEADIKG EWLEGRNLAS AAVYRARKNN  
551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR  
601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRTIGAGVR RQETHTDPA  
651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVDNLFNKH  
701 YRTQPDHSY GALRTVNAAF TYRFFK\*

25 ORF23ng-1 (SEQ ID NO: 672) and ORF23-1 (SEQ ID NO: 666) show 95.9% identity in 725 aa overlap:

30  
35  
40  
45  
50

	10	20	30	40	50	60
orf23-1.pep	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
orf23ng-1	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
	10	20	30	40	50	60
orf23-1.pep	70	80	90	100	110	120
orf23ng-1	PLGLPMTLRE	IPQSVSVITS	QQMRDQNIKT	LDRALLQATG	TSRQIYGSDR	AGYNYLFARG
	70	80	90	100	110	120
orf23-1.pep	130	140	150	160	170	180
orf23ng-1	SRIANYQING	IPVADALADT	GNANTAAYER	VEVVRGVAGL	PDGTGEP SAT	VNLVRKHPTR
	130	140	150	160	170	180
orf23-1.pep	190	200	210	220	230	240
orf23ng-1	KPLFEVRAEA	GNRKHFGLD	ADVSGSLNTE	GTLRGRLVST	FGRGDSWR	RRERSRDAELYGI
	190	200	210	220	230	240
orf23-1.pep	250	260	270	280	290	300
orf23ng-1	LEYDIAPQTR	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWANSRHRAL
	250	260	270	280	290	300

-487-

		310	320	330	340	350	360
	orf23-1.pep	NLFAGIEHFRNQDWKLKA	EYDYTRSRFRQPYGV	AGVLSIDHNTAATDLIP	GYWHADPRTH		
	orf23ng-1	NLFAGIEHFRNQDWKLKA	EYDYTRSRFRQPYGV	AGVLSIDHNTAATDLIP	GYWHADPRTH		
5		310	320	330	340	350	360
	orf23-1.pep	SASVSLIGKYRLFGREHD	LIAGINGYKYASNKYGER	SIIPNAIPNAYEFSRTG	AYPQPAS		
	orf23ng-1	SASMSLTGKYRLFGREHD	LIAGINGYKYASNKYGER	SIIPNAIPNAYEFSRTG	AYPQPSS		
10		370	380	390	400	410	420
	orf23-1.pep	FAQTIPQYGTTRRQIGGY	LATRFRAADNLSLILGG	RYTRYRTGSYDSRTQG	MTYVSANRFT		
	orf23ng-1	FAQTIPQYDTRRQIGGY	LATRFRAADNLSLILGG	RYTRYRTGSYDSRTQG	MTYVSANRFT		
15		430	440	450	460	470	480
	orf23-1.pep	PYTGIVFDLTGNLSLYGS	YSSSLFVPQSQKDEHGS	YLPVGTGNNLEAGIKGE	WLEGRNLAS		
	orf23ng-1	PYTGIVFDLTGNLSLYGS	YSSSLFVPQQLQKDEHGS	YLPVGTGNNLEADIKGE	WLEGRNLAS		
20		490	500	510	520	530	540
	orf23-1.pep	AAVYRARKNNLATAAGRD	PSGNTYYRAANQAKTHG	WEIEVGGRTPEWQIQAG	YSQSKTR		
	orf23ng-1	AAVYRARKNNLATAAGRD	QSGNTYYRAANQAKTHG	WEIEVGGRTPEWQIQAG	YSQSKPR		
25		550	560	570	580	590	600
	orf23-1.pep	DQDGSRLNPDSVPERSFK	LFTAYHFAPEAPSGWTI	GAGVRWQSEHTDPATLR	IPNPAAK		
	orf23ng-1	DQDGSRLNPDSVPERSFK	LFTAYHLAPEAPSGRTI	GAGVRRQGETHTDPAAL	RIPNPAAK		
30		610	620	630	640	650	660
	orf23-1.pep	ARAADNSRQKAYAVADIM	ARYRFNPRAELSLNVDN	LFNKHYRTQPDHRHSYG	ALRTVNAAF		
	orf23ng-1	ARAVANSRQKAYAVADIM	ARYRFNPRTLSLNVN	LFNKHYRTQPDHRHSYG	ALRTVNAAF		
35		670	680	690	700	710	720
	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					
40							

In addition, ORF23ng-1 (SEQ ID NO: 672) shows significant homology with an OMP (SEQ ID NO: 1155) from *E.coli*:

sp|P16869|FHUE\_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR )gi|1651542|gnl|PID|d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli] )gi|1651545|gnl|PID|d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] )gi|1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729  
Score = 332 bits (843), Expect = 3e-90

Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)

Query: 38 TITVTADRTASSN--DGYTVSGHTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRL 95  
T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL +  
Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSGATKMQMTQORDIPQSVTIVSQRMEDQQLQTLGEVM 102

5 Query: 96 LQATGTSRQIYGSDRAGYNYLFARGSRANYQINGIP-----VADALADTGNANTAA 147  
G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A  
Sbjct: 103 ENTLGISKSQADSDRALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154

10 Query: 148 YERVEVVRGVAGLPDGTGEPSATVNLVRKHPTKPLF-EVRAEAGNRKHFGLGADVSGSL 206  
+ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L  
Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSPL 214

Query: 207 NAEGTLRGRLVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADA 266  
+G +R R+V + DSW S GI++ D+ T + AG +YQ+ +  
Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDRYNSEKTFFGSIVDADLGLDLTLSAGYEYQRIDVNSPT 274

15 Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSRNRALNLFAGIEHRFNQDWKLKAEYDYTRSR 326  
+++ G + ++ + A +W+ + +F ++ +F W+ ++  
Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334

Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLFG 374  
F + Y A V D ++ PG+ W++ R A + G Y LFG  
Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394

20 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432  
R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT  
Sbjct: 395 RQHNLMFQ-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451

Query: 433 QIGGYLATRFRAADNLSLILGGYRSRYRAGSYNSRTQGMY-VSANRFTPYTGIVFDXXX 491  
Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD  
25 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504

Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPVTGNLEADIKGEWLEGRNLASAAVYRARKNNL 551  
F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+  
Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAIPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564

Query: 552 ATAAGR--DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608  
A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N  
30 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624

Query: 609 PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSR 668  
P ++P + K+FT+Y L P P T+G GV Q +TD P RA  
Sbjct: 625 P-NLPRTTVKMFTSYRL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGTFR-----E 672

35 Query: 669 QKAYAVADIMARYRFPNRTLSLNVNLFNKHYRTQPDRL-SYGALRTVNAAFTYRF 724  
Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F  
Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTNRNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 ORF23-1 (SEQ ID NO: 666) (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by

SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 (SEQ ID NO: 666) is a surface-exposed protein, and that it is a useful immunogen.

### Example 80

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 673):

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
151 AGCGTCAGcA CGCCTGCTTC GGCGGcGgCa ATCATACCTT CGTCTTCGGA
201 AACGGGGATA AACGcGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TGCGTGCCAC
15  351 TnAGTCGCCG ACGGGG..

```

This corresponds to the amino acid sequence (SEQ ID NO: 674; ORF24):

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
20  101 PCVPQTLKPI XSRMRATXSP TG..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 675):

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
25  101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
151 AGCGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
201 AACGGGGATA AACGCGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TCTTCAAGAA TGCGTGCCAC
30  351 TGAGTCGCCG ACGGCGGGGG TCGGCGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCCACG
451 CGGGTAATTT TGAAAGCAGT TTTCTTCACT ACTTCCGCAA CTTCGGTCAA
501 TGTCGTTGCA TCTGAATTTT CCAACGCGGC TTTTACGACA CCTGGGCCGG
551 ATACGCCGAC ATTGATAACG GCATCCGCTT CGCCCGAACC ATGAAACGCG
35  601 CCCGCCATAA ACGGGTTGTC TTCCACCGCG TTGCAGAACA CGACAATTTT
651 AGCGCAGCCG AAACCTTCGG GCGTGATTTC CGCCGTGCGT TTGACGGTTT
701 CGCCCGCCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTCGGGAAT
801 GGAGCGGATT AACACCTCAT CCGAAGGCGA CATCCCTTTT TGCACCAACG
40  851 CGGAAAAACC GCCGATAAAA GACACACCGA TGGCTTTGGC AGCTTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 676; ORF24-1):

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS

```

-490-

51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV  
 101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT  
 151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP\*NA  
 201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLP  
 251 ILMELHTISV VFIASGMERI NTSSEGDIPF CTNAEKPPIK DTPMALAALS  
 301 KVCATLT\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF24 (SEQ ID NO: 674) shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) (SEQ ID NO: 678) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf24a.pep	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA					
15	orf24	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTAQAVMASSLSNVSTPASAAA					
		10	20	30	40	50	60
	orf24a.pep	IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP					
20	orf24	IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP					
		70	80	90	100	110	120
	orf24a.pep	IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP					
25	orf24	IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP					
		70	80	90	100	110	120
	orf24a.pep	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT					
30	orf24	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT					
		130	140	150	160	170	180
	orf24a.pep	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT					
35	orf24	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT					
		130	140	150	160	170	180
	orf24a.pep	PGPDTPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA					
40	orf24	PGPDTPTLITASASPEPXNAPAIINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
		190	200	210	220	230	240
	orf24a.pep	PGPDTPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA					
45	orf24	PGPDTPTLITASASPEPXNAPAIINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
		190	200	210	220	230	240
	orf24a.pep	SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS					
50	orf24	SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
	orf24a.pep	SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS					
55	orf24	SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
	orf24a.pep	KVCATLT					
60	orf24	KVCATLT					

The complete length ORF24a nucleotide sequence (SEQ ID NO: 677) is:

1 ATGCGCACGG CAGTGGTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC  
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA  
 45 101 TCATATCCAA NCCGACCGAA CAAACGCGG TCATCGCTTC GAGTTTATCC



5  
10  
15

```

151 AACGTCAGCA CGCCTGCTTC GCGCGCGGCA ATCATACCTT CGTCTTCGGA
201 NACGGGGATA AACGCGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA
251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGCGCGCCAC
351 CGAGTCGCCG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCCACG
451 CGGGTAATTT TGAAGGCGGT TTTCTTCACA ACTTCGGCAA CTTCGGTCAA
501 TGTCGTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG
551 ATACGCGGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
601 CCCGCCATAN ACGGGTTGTC TTCCNCCGCG TTGCAGAACA CGACGATTTT
651 GGCGCAGCCG AAACCTTCTA GTGTGATTTT ANCCGTGCGT TTGATGGTTT
701 CGCCCGCCAG TCTGACCGCG TCCATATTGA TACCGGCGCG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
801 GGAACGGATN AACACCTCGT CAGAAGGCGA CATACCTTTT TGCACCAGCG
851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 678):

20  
25

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51  NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDPTLIT ASASPEP*NA
201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
251 ILMELHTISV VFASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
301 KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a (SEQ ID NO: 678) and ORF24-1 (SEQ ID NO: 676) show 96.4% identity in 307 aa overlap:

30  
35  
40  
45

```

      10      20      30      40      50      60
orf24a.pep  MRTAVVLLLIMPMAASSAMMPMVCGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
orf24-1      MRTAVVLLLIMPMAASSAMMPMVCGVSPGTAIISKPTEQTAVMASSLSVSTPASAAA
      10      20      30      40      50      60

      70      80      90     100     110     120
orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPFFFTASFSNAKAAVPCVPQTLKPISSRMRATESP
orf24-1      IIPSSSETGINAPLKPPTALEAIMPFFFTASFSNAKAAVPCVPQTLKPISSRMRATESP
      70      80      90     100     110     120

      130     140     150     160     170     180
orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
orf24-1      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      130     140     150     160     170     180

      190     200     210     220     230     240
orf24a.pep  PGPDPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
orf24-1      PGPDPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
      190     200     210     220     230     240

```

		250	260	270	280	290	300
orf24a.pep		SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS					
orf24-1		SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
5		250	260	270	280	290	300
	orf24a.pep	KVCATLTX					
	orf24-1	KVCATLTX					

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 (SEQ ID NO: 674) shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) (SEQ ID NO: 680) from *N.gonorrhoeae*:

	orf24.pep	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA	60
15	orf24ng	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTEQAVMASSLSVNTPASAAA	60
	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFTSASFSNAKAAVPCVPQTLKPIXSRRMRATXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFTSASFSNAKAAVPCVPQTLKPISSRRMRATESP	120
	orf24.pep	TG	122
20		:	
	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVRILTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence (SEQ ID NO: 679) is:

	1	ATGCGCACGG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
25	51	GGCGATGATG	CCGGAATG	TGTGCGCGGG	CGTGTGCGCCG	GGAACGGCAA
	101	TCATGTCCAA	ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTC
	151	AGCGTCAACA	CGCCTGCCTC	GGCGGCGGCA	ATCATACCTT	CGTCTTCGGA
	201	AACGGGGATA	AACGCGCCGC	TCAAACCGCC	GACCGCGCTG	GAAGCCATCA
	251	TGCCGCCCTT	TTTACGGCA	TCGTTACGCA	ATGCCAAAGC	TGCTGTTGTG
30	301	CCGTGCGTAC	CGCAGACGCT	CAAGCCCAT	TCTTCAAGAA	TGCGCGCCAC
	351	CGAGTCGCGG	ACGGCGGGGG	TCGGTGCCAG	CGACAAATCG	AGAATGCCGA
	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GACCGATGAG	TTCGCCCACG
	451	CGGGTGATTT	TGAAAGCGGT	TTCTTTCACG	ACTTCGGCGA	CCTCGGTCAG
	501	GCTGACCGCG	TCCGAATTTT	CCAGCGCGGC	TTTGACCACG	CCTGGACCGG
35	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCCGAGCC	GTGGAACGCA
	601	CCCGCCATAA	ACGGATTGTC	TTCCACCGCG	TGCGAGAACA	CGACGATTTT
	651	GGCGCAGCCG	AAACCTTCGG	GTGTGATTTC	AGCCGTGCGT	TTGATGGTTT
	701	CGCCTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG	CGTGTGCCC
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTCATCG	CTTCGGGAAC
40	801	GGAACGGATC	AACACCTCAT	CCGAAGGCGA	CATACCTTTT	TGCACCAGCG
	851	CGGAAAAGCC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC	TGCCTTGTCC
	901	AAAGTCTGCG	CCACGCTGAC	ATAA		

This encodes a protein having amino acid sequence (SEQ ID NO: 680):

45	1	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
	51	SVNTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPFFTA	SFSNAKAAV
	101	PCVPQTLKPI	SSRRMRATESP	TAGVGASDKS	RMPNGIFSIF	EASRPMSSPT

151 RVILKAVFFT TSATSVRLTA SEFSSAALTT PGPDTPTLIT ASASPEPWNA  
 201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LMVSPASLTA SILIPARVLP  
 251 ILMELHTISV VFIASGTERI NTSSEGDIPF CTSAEKPPIK DTPMALAALS  
 301 KVCATLT\*

5

ORF24ng (SEQ ID NO: 680) and ORF24-1 (SEQ ID NO: 676) show 96.1% identity in 307 aa overlap:

10	orf24-1.pep	10 20 30 40 50 60	MRTAVVLLLLIMPMMAASSAMPEMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA 
	orf24ng	10 20 30 40 50 60	MRTAVVLLLLIMPMMAASSAMPEMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA 
15	orf24-1.pep	70 80 90 100 110 120	IIPSSSETGINAPLKPTALEAIMPPFFTASFNAKAADVPCVPQTLKPISSRMATESP 
	orf24ng	70 80 90 100 110 120	IIPSSSETGINAPLKPTALEAIMPPFFTASFNAKAADVPCVPQTLKPISSRMATESP 
20	orf24-1.pep	130 140 150 160 170 180	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAFFT 
	orf24ng	130 140 150 160 170 180	TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVRLTASEFSSAALTT 
25	orf24-1.pep	190 200 210 220 230 240	PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA 
	orf24ng	190 200 210 220 230 240	PGPDTPTLITASASPEPWNAIPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA 
30	orf24-1.pep	250 260 270 280 290 300	SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS 
	orf24ng	250 260 270 280 290 300	SILIPARVLPILMELHTISVVFIASGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS 
35	orf24-1.pep		KVCATLT 
	orf24ng		KVCATLT

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 681):

```

1  ..ACCGACGTGC AAAAAGAGTT GGTCCGCGAA CAACGCAAGT GGGCGCAGGA
51  AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
101 ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
151 ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 682; ORF25):

```

1  ..TDVQKELVGE QRKWAQEKIS NCRQAAAQAD RQEYAEYLKL QCDTRMTRE
51  IQYLRGYSID *

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 683):

```

1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
101 TGCAAGGCAT ACGCGGCAAT ATTCAGGAAA CGCTCACGCA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG AAAAAATTAT
201 CGCCGCCGCC TACGTTTGG CGTTTCTTT GGAACACGCT TCGGAAACGC
251 AGGAAGGCGG GCGCACGTTC TGTATCGCCG ATTTGAACAT TACCGTGCCG
301 TCTGAAACGC TTGCCGATGC CAAGGCAAC AGCCCCTGT TGTACGGGGA
351 AACTGCTTTG TCGGATATTG TGCGGCAGAA GACGGGCGGC AATGTGAGT
401 TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCCTGCC CGTCAAAGAC
451 GGTGACACGG CATTTGTGCA CAACACGGTC GGTATGGCGG CGCAAACGCT
501 GTCTGCCGCG CTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT GAAAAAAGAA GACGCGGTCA GGATTTTGAG CGGAAAAGCC
601 CGTGAAGAAG AACCGTCCAA ACCCACGCC GAAGACATTT TGGAACACAA
651 TGCCGCCGGC GGCATGCGG GCGTACCCCA AGCCGCAGAA GGCGCGCCCG
701 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
751 GTATCACGGG GCGAAGTGGA AGAGCGCGC GTACAAAACC AGCGTGCGGA
801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGGTCGG CGAACACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
1001 GCTATTCCAT CGATTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 684; ORF25-1):

```

1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
51  SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQTAFVDNTV GMAAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEESPSTPT EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLQC DTRMTREIRI YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 (SEQ ID NO: 682) shows 98.3% identity over a 60aa overlap with an ORF (ORF25a)

(SEQ ID NO: 686) from strain A of *N. meningitidis*:

-495-

```

      orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLDQKELVGEXRKWAQEKISNCRQAAAQAD
                  250      260      270      280      290      300
5      orf25.pep      RQEYAEYLKLQCDTRMRERIQYLRGYSIDX
                        |||||
      orf25a      RQEYAEYLKLQCDTRMRERIQYLRGYSIDX
                  310      320      330

```

10 The complete length ORF25a nucleotide sequence (SEQ ID NO: 685) is:

```

      1  ATGTATCGGA AACTCATGTC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
      51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
     101  TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCACGCA GGAAGCGCGT
     151  TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
     15  201  CGCCGCCGCC TANGTNNNGN NGTNTCTTT GGAACACGCT TCGGAAACGC
     251  AGGAAGGCGG GCGCACGTTT TGTNTCGCCG ATTTGAACAT TACCGTGCCG
     301  TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGC TGTACGGGGA
     351  AACCGCTTTG TCGGATATTG TCGGCGAGAA GACGGGCGGC AATGTCGAGT
     401  TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCCTACC CGTCAAAGAC
     20  451  GGTCAGANGG CATTTGTCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
     501  GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
     551  GCAAGGCGGT AAAAAAAGAA GACGCGGTCA GGATNTGAG CNGANAAGCC
     601  CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATT TGGAAACATA
     651  TGCCGCCGGA GGGGATGCAG ACGTACCCCA AGCCGGAGAA GACGCGCCCG
     25  701  AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
     751  GTATCACGGG GCGAAGTGGA AGAGGCGCGN GTACAAAACC AGCGTGCGGA
     801  ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
     851  AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
     901  CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
     30  951  GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
    1001 GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 686):

```

     35  1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
      51  SFAREDXXQF VDADXIIAAA XXXXXSLEHA SETQEGGRTF CXADLNITVP
     101  SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
     151  GQXAFVDNTV GMAAQTL SAA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA
     201  REXEPSKXXP EDILEHNAAG GDADVQAGE DAPEPEILHP DDGERADTVT
     251  VSRGEVEEAR VQNQRAESEI TKLWGGLDQD VQKELVGEXR KWAQEKISNC
     40  301  RQAAAQADRQ EYAEYLKLQC DTRMRERIQ YLRGYSID*

```

ORF25a (SEQ ID NO: 686) and ORF25-1 (SEQ ID NO: 684) show 93.5% identity in 338 aa overlap:

```

      45      orf25a.pep      MYRKLIALPFALLLAACGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
                        |||||
      orf25-1      MYRKLIALPFALLLAACGRE EPPKALECANPAVLQGIRGNIQETLTQEARSFAREDGRQF
                        |||||
      50      orf25a.pep      VDADXIIAAXXXXSLEHASETQEGGRTFCXADLNITVPSETLADAKANSPLLYGETAL
                        |||||

```

5	orf25-1	VDADKIIIAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL	70	80	90	100	110	120
	orf25a.pep	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQXAFVDNTVGMAAQTLAALLPYGVKSIV	130	140	150	160	170	180
10	orf25-1	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQTAFVDNTVGMAAQTLAALLPYGVKSIV	130	140	150	160	170	180
	orf25a.pep	MIDGKAVKKEDAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVPQAGEDAPEPEILHP	190	200	210	220	230	240
15	orf25-1	MIDGKAVKKEDAVRILSGKAREEPEPSKPTPEDILEHNAAGGDAGVPPQAEGAPEPEILHP	190	200	210	220	230	240
	orf25a.pep	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNC	250	260	270	280	290	300
20	orf25-1	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNC	250	260	270	280	290	300
	orf25a.pep	RQAAAQADRQEYAEYLKLQCDTRMTRERIQYLRGYSID	310	320	330	339		
	orf25-1	RQAAAQADRQEYAEYLKLQCDTRMTRERIQYLRGYSID	310	320	330			

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 (SEQ ID NO: 682) shows 100% identity over a 60aa overlap with a predicted ORF  
 25 (ORF25ng) (SEQ ID NO: 688) from *N.gonorrhoeae*:

30	orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
	orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNCRQAAAQAD	308
30	orf25.pep	RQEYAEYLKLQCDTRMTRERIQYLRGYSID	60
	orf25ng	RQEYAEYLKLQCDTRMTRERIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence (SEQ ID NO: 687) is:

35	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCGGTG
40	101	TGCAGGACAT	ACGCGGCAGT	ATTGAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
45	201	CGCCGCCGCC	TACGGTTTGG	CGTTTCTTT	GGAACACGCT	TCGGAACGC
	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
50	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
55	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCTTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
60	501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTGTAG	CGGCAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCC	GAAGACATTT	TGGAACACAA

651 TGCCGCCGGC GGCGATGCGG GCGTACCCCA AGCCGCAGAA GGCGCACCCG  
701 AACCCGAAAT CCTGCATCCC GACGACGTCG AGCGTGCCGA TACCGTTACC  
751 GTATCACGGG GCGAAGTGA AGAGGCGCGC GTACAAAACC AACGTGCGGA  
801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG  
851 AGTTGGTCGG CGAACAGCGC AAGTGGGCGC AGGAAAAAAT CAGcaactgc  
901 cgACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA  
951 GCTCCAATGC GACACGCGGA TGACGCGCGA ACggaTACAG TATCTTCGCG  
1001 GCTATTCCAT CGATTAG

10 This encodes a protein having amino acid sequence (SEQ ID NO: 688):

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR  
51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP  
101 SETLADAEAN SPALLYGETSL ADIVQQKTGG NVEFKDGVLT AAVRFLPAKD  
151 ARTAFIDNTV GMATQTLASA LLPYGVKSIV MIDGKAVTKE DAVRVLSGKA  
201 REEESKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT  
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC  
301 RQAAAQADRQ EYAEYKLQCD TRMTRERIQ YLRGYSID\*

ORF25ng (SEQ ID NO: 688) and ORF25-1 (SEQ ID NO: 684) show 95.9% identity in 338 aa  
20 overlap:

		10	20	30	40	50	60
orf25-1.pep	MYRKLIALPF	FALLLAACGRE	EPPKALECAN	PAVLQDIRGS	IQETLTQEAR	SFAREDGRQF	
orf25ng	MYRKLIALPF	FALLLAACGRE	EPPKALECAN	PAVLQDIRGS	IQETLTQEAR	SFAREDGRQF	
		10	20	30	40	50	60
orf25-1.pep	VDADKIIAAA	YGLAFSLEHA	SETQEGGRTF	CIADLNITVP	SETLADAEAN	SPALLYGETAL	
orf25ng	VDADKIIAAA	YGLAFSLEHA	SETQEGGRTF	CIADLNITVP	SETLADAEAN	SPALLYGETSL	
		70	80	90	100	110	120
orf25-1.pep	SDIVRQKTGG	NVEFKDGVLT	AAVRFLPVK	DGQTAFVDNT	VGMAAQTLSA	ALLPYGVKSIV	
orf25ng	ADIVQQKTGG	NVEFKDGVLT	AAVRFLPAK	DARTAFIDNT	VGMAATQTLA	ALLPYGVKSIV	
		130	140	150	160	170	180
orf25-1.pep	MIDGKAVKKED	AVRILSGKARE	EEPSKPTPED	ILEHNAAGGD	AGVPQAAEGA	PEPEILHP	
orf25ng	MIDGKAVTKED	AVRVLSGKARE	EEPSKPTPED	ILEHNAAGGD	AGVPQAAEGA	PEPEILHP	
		190	200	210	220	230	240
orf25-1.pep	DDGERADTVT	VSRGEVEEAR	VQNQRAESEI	TKLWGGLDTD	VQKELVGEQR	KWAQEKISNC	
orf25ng	DDVERADTVT	VSRGEVEEAR	VQNQRAESEI	TKLWGGLDTD	VQKELVGEQR	KWAQEKISNC	
		250	260	270	280	290	300
orf25-1.pep	RQAAAQADRQ	EYAEYKLQCD	TRMTRERIQ	YLRGYSID	X		
orf25ng	RQAAAQADRQ	EYAEYKLQCD	TRMTRERIQ	YLRGYSID	X		
		310	320	330	339		

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 5 ORF25-1 (SEQ ID NO: 684) (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive  
10 result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 (SEQ ID NO: 684) is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1 (SEQ ID NO: 684).

### Example 82

- 15 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 689)

```

      1  ATGCAGCTGA  TCGACTATTC  ACATTCATTT  TTCTCGGTTG  TGCCACCCTT
      51  TTTGGCACTG  GCACTTGCCG  TCATTACCCG  CCGCGTACTG  CTGTCTTTAG
     101  GCATCGGTAT  TCTGGwysGC  GTTGCCTTTT  TGGTCGGCGG  CAACCCCGTC
     151  GACGGTCTGA  CACACCTGAA  AGACATGGTC  GTCGGCTTGG  CTTGGTCAGA
     201  CGsyGATTGG  TCGCTGGGCA  AACCAAAAT  CTTGGTTTTT  CkGATACTTT
     251  TGGGTATTTT  TACTTCCCTG  CTGACCTACT  CCGGCAGCAA  T.....

                               //

     851  .....AC  TTCGCTGGTA
     901  TTCGGCGGCA  CTTGCGGCGT  CTTTGCCGTC  GTTCTCTGCA  CGCTCGGCAC
     951  GATTAAAACC  GCCGACTATC  CCAAAGCCGT  TTGGCAGGGT  GCGAAATCTA
    1001  TGTTCCGGCGC  AATCGCCATT  TTAATCCTCG  CTTGGCTCAT  CAGTACGGTT
    1051  GTCGGCGAAA  TGCACACCGG  CGATTACCTC  TCCACACTGG  TTGCGGGCAA
    1101  CATCCATCCC  GGCTTCCTGC  CCGTCATCCT  CTTCTGCTC  GCCAGCGTGA
    1151  TGGCGTTTGC  CACAGGCACA  AGCTGGGGGA  CGTTCGGCAT  TATGCTGCCG
    1201  ATTGCCGCCG  CCATGGCGGT  CAAAGTCGAA  CCCGCGCTGA  TTATCCCGTG
    1251  TATGTCCGCA  GTAATGGCGG  GGGCGGTATG  CGGCGACCAC  TGCTCGCCCA
    1301  TTTCCGACAC  GACCATCCTG  TCGTCCACCG  GCGCGCGCTG  CAACCACATC
    1351  GACCACGTTA  CCTCGCAACT  GCCTTACGCC  TTAACCGTTG  CCGCCGCCGC
    1401  CGCATCGGGC  TACCTCGCAT  TGGGTCTGAC  AAAATCCGCG  CTGTGGGGCT
    1451  TTGGCAGGAC  AGGCATTGTA  TTGGCGGTGC  TGATTTTCT  GTTGAAAGAT
    1501  AAAAAA..

```

This corresponds to the amino acid sequence (SEQ ID NO: 690; ORF26):



1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFVLVGGNPV  
51 DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...

//

251 .....TSLV  
301 FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
351 VGEMHTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH  
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
501 KK..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 691):

1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT  
51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG  
101 GCATCGGTAT TCTGGTCGGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC  
151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA  
201 CGGCGATTGG TCGCTGGGCA AACCAGAAAT CTTGGTTTTT CTGATACTTT  
251 TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT  
301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAAATGCT  
351 GACCGCCTGC CTCGTGTTCG TAACCTTTAT CGACGACTAT TTCCACAGTC  
401 TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC  
451 CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT  
501 GCTGATGCCC GTTTCAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG  
551 GACTGCTCGT TACCTACAAA ATCACCAGAT ACACGCCGAT GGGGACGTTT  
601 GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT  
25 651 GTTCGTCGTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTCG  
701 AACAAAGCCG GTTGAACGAA GCCCAGATG AAAGTGCCTT TTCAGACGCT  
751 ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC  
801 CTCACCGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT  
851 TCAGCATTTT GGGGGCATTT GAAAACACGG ACGTAAACAC TTCGCTGGTA  
30 901 TTCGGCGGCA CTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC  
951 GATTAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA  
1001 TGTTCCGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT  
1051 GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA  
1101 CATCCATCCC GGCTTCTGTC CCGTCATCCT CTTCTGCTC GCCAGCGTGA  
35 1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG  
1201 ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG  
1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA  
1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC  
1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
40 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT  
1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT  
1501 AAAAAACGCG CCAACGCCTG A

This corresponds to the amino acid sequence (SEQ ID NO: 692; ORF26-1):

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV  
51 DGLTHLKDMV VGLAWSDDGW SLGKPKILVF LILLGIFTSL LTYSGSNQAF  
101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAIA RPVTDKFKVS  
151 RTKLAYILDS TAAPMCVLMV VSSWGASIIA TLAGLLVYK ITEYTPMGTF  
201 VAMSLMNYA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAVSDA  
251 TKGRVYALII PVLALIASTV SAMIYGAQA SETFSILGAF ENTDVNTSLV  
301 FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
351 VGEMHTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH  
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
55 501 KKRANA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 (SEQ ID NO: 1156) of *H.influenzae* (accession number P44263)

ORF26 (SEQ ID NO: 690) and HI1586 (SEQ ID NO: 1156) show 53% and 49% amino acid identity in 97 and 221 aa overlap at the N-terminus and C-terminus, respectively:

```

Orf26   1  MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60
          M+LID+S S +S+VP LA+ LA+ TRRV L +L V
HI1586  14  MELIDFSSSVWSIVPALLAIILAIATRRVLVLSAGIIGSLMLSQWIGSAFNILVKNV 73

Orf26   61  VGLAWSDXDWSLGKPKILVFXILLGIFTSLITYSGSN 97
          V L ++D + + I++F +LLG+ T+LLT SGSN
HI1586  74  VSLVYADGEIN-SNMNIVLFLLLLGLVLTALLTVSGSN 109

//

Orf26   86  IFTSLITYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGAQSMFGXXXX 141
          +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
HI1586  299 VFSVLGTFENTVVGTSVLVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAI 358

Orf26   142 XXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP 201
          + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP
HI1586  359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLP 418

Orf26   202 IAAAMAVKVEPALIIPCSAVMAGAVCGDHCSPISTTILSSTGARNHIDHVTQSXXXX 261
          IAAAMA P L++PC+SAVMAGAVCGDHCSPISTTILSSTGA+CNHIDHVT+Q
HI1586  419 IAAAMAANAPELLLPCLSAVMAGAVCGDHCSPISTTILSSTGAKCNHIDHVTQLPYA 478

Orf26   262 XXXXXXXXXXXXXXXXXXXKSALLGFGTTGIVLAVLIFLLKDK 302
          S L GF T + L V+IF +K +
HI1586  479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLIVIIFAVKKR 519

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF26 (SEQ ID NO: 690) shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) (SEQ ID NO: 694) from strain A of *N. meningitidis*:

```

          10      20      30      40      50      60
orf26.pep MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV
          |||
orf26a    MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
          10      20      30      40      50      60

          70      80      90      99
orf26.pep VGLAWSDXDWSLGKPKILVFXILLGIFTSLITYSGSNXX-----
          |||
orf26a    VGLAWSDXDWSLGKPKILVFXILLGIFTSLITYSGSNQAFADWAKRHIKNRRGAKMLTAC
          70      80      90      100     110     120

```

5	orf26.pep	-----
	orf26a	<u>LVFVTFIDDFHSLAVGAXARPVTDKFKVSRKLAYILDSTAAPMCVLMPVSSWGASIIA</u> 130 140 150 160 170 180
10	orf26.pep	-----
	orf26a	<u>TLAGLLVTYKITEYTPMGTFVAMSLMNYIALFALIMVFVVAWFSDIGSMARFEQAALNE</u> 190 200 210 220 230 240
15	orf26.pep	-----100 110-----TSLV 
	orf26a	<u>AHDETAVSDGSWGRVYALIIPVLALIASTVSAMIYTGAAQASETFSILGAFENTDVNTSLV</u> 250 260 270 280 290 300
20	orf26.pep	120 130 140 150 160 170 <u>FGGTCGVFAVVLCTLTGTIKTADYPKAVWQGA</u>      :
	orf26a	<u>FGGTCGVLA VVLCTLTGTIKIADYPKAVWQGA</u> 310 320 330 340 350 360
25	orf26.pep	180 190 200 210 220 230 <u>STLVAGNIHPGFLPVILFLLASVMAFATGTSWGT</u>      :
	orf26a	<u>STLVAGNIHPGFLXVILFLLASVMAFATGTSWGT</u> 370 380 390 400 410 420
30	orf26.pep	240 250 260 270 280 290 <u>VMAGAVCGDHCSPISD</u>      :
	orf26a	<u>VMAGAVCGDHCSPISD</u> 430 440 450 460 470 480
35	orf26.pep	300 310 <u>LLGFGTTGIVLAVLIFLLKDKK</u>      :
	orf26a	<u>LLGFGXTGIVLAVLIFLLKDKK</u> 490 500

The complete length ORF26a nucleotide sequence (SEQ ID NO: 693) is:

40	1	ATGCAGCTGA	TCGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
	51	TTTGGCACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
45	101	GCATCGGTAT	TCTGGTCGGC	GTTGCCTTTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGTCAGA
50	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTC	CTGATACTTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
45	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGGCGCGGCG	CGAAAATGCT
	351	GACCGCCTGC	CTCGTGTTTC	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
50	401	TCGCCGTCGG	TGCGNTTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGGCCCAAAC	TCGCCTACAT	CCTCGACTCC	ACTGCCGCGC	CTATGTGCGT
50	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
50	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGGT
	651	GTTTCGTCGTC	GCATGGTTCT	CCTTCGACAT	CGGCTCGATG	GCACGTTTCG
	701	AACAAGCCGC	GTTGAACGAA	GCCCACGATG	AAACTGCCGT	TTCAGACGGC

5  
10  
15

```

751 AGCTGGGGCA GGGTTTACGC ATTGATTATT CCCGTTTGG CCTTAATCGC
801 CTCAACGGTT TCCGCCATGA TCTACACCGG TGCACAGGCA AGCGAAACCT
851 TCAGCATTTT GGGTGCATTT GAAAATACGG ACGTGAACAC TTCGCTGGTA
901 TTCGGCGGCA CTTGCGGCGT GCTTGCCGTC GTCCTCTGCA CGCTCGGCAC
951 GATTAATAATC GCCGATTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCCA
1001 TGTTCCGGCGC AATCGCCATT TTAATCCTTG CCTGGCTCAT CAGTACGGTT
1051 GTCGGCGAAA TGCACACAGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
1101 CATCCATCCC GGCTTCCTGN CCGTCATCCT TTTCTGCTC GCCAGCGTGA
1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT CATGCTGCCG
1201 ATTGCCGCGG CCATGGCGGT CAAAGTCGAT CCCTCACTGA TTATCCCGTG
1251 TATGTCCGCC GTGATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT
1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
1501 AAAAAACGCG CCAACGCCTG A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 694):

20  
25  
30

```

1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
51  DGLTHLKDMV VGLAWSGDGW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF
101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAXA RPYTDKFKVS
151 RAKLAYILDS TAAPMCVLMP VSSWGASIIA TLAGLLVITYK ITEYTPMGTF
201 VAMSLMNYA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAUSDG
251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
301 FGGTCGVLAV VLCTLGTIKI ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEMHTGDYL STLVAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLFGXTGIV LAVLIFLLKD
501 KKRANA*

```

ORF26a (SEQ ID NO: 694) and ORF26-1 (SEQ ID NO: 692) show 97.8% identity in 506 aa overlap:

35  
40  
45  
50

```

              10      20      30      40      50      60
orf26a.pep    MQLIDYSHSF FSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
              |||||||
orf26-1       MQLIDYSHSF FSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
              10      20      30      40      50      60

              70      80      90      100     110     120
orf26a.pep    VGLAWSGDGWSLGKPKXLVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
              |||||||
orf26-1       VGLAWSGDGWSLGKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
              70      80      90      100     110     120

              130     140     150     160     170     180
orf26a.pep    LVFVTFIDYFHSLAVGAXARPYTDKFKVSRKLAYILDSTAAPMCVLMPVSSWGASIIA
              |||||||
orf26-1       LVFVTFIDYFHSLAVGAIARPYTDKFKVSRKLAYILDSTAAPMCVLMPVSSWGASIIA
              130     140     150     160     170     180

              190     200     210     220     230     240
orf26a.pep    TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE
              |||||||
orf26-1       TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE
              190     200     210     220     230     240

```

		250	260	270	280	290	300
	orf26a.pep	AHDETA	VDGSGW	GRVYAL	IIPVLAL	IASTVS	AMIYTGAQ
	orf26-1	AHDETA	VDGSGW	GRVYAL	IIPVLAL	IASTVS	AMIYTGAQ
5		250	260	270	280	290	300
	orf26a.pep	FGGTCG	VLAVVL	CTLGTI	KIADYP	KAVWQG	AKSMFGA
	orf26-1	FGGTCG	VLAVVL	CTLGTI	KIADYP	KAVWQG	AKSMFGA
10		310	320	330	340	350	360
	orf26a.pep	STLVAG	NIHPGF	LXVILF	LLASVM	AFATGT	SWGTFG
	orf26-1	STLVAG	NIHPGF	LXVILF	LLASVM	AFATGT	SWGTFG
15		370	380	390	400	410	420
	orf26a.pep	IMLP	IAAAM	AVKV	DP	SLI	IPCMSA
	orf26-1	IMLP	IAAAM	AVKV	DP	SLI	IPCMSA
		370	380	390	400	410	420
	orf26a.pep	VMAGAV	CGDHCS	PISD	TTILS	STGARC	NHIDHVT
	orf26-1	VMAGAV	CGDHCS	PISD	TTILS	STGARC	NHIDHVT
20		430	440	450	460	470	480
	orf26a.pep	SQLPY	ALTVA	AAAA	ASGY	LALGL	TKSA
	orf26-1	SQLPY	ALTVA	AAAA	ASGY	LALGL	TKSA
		430	440	450	460	470	480
	orf26a.pep	LLGFG	XTGIV	LAVLI	FLLK	DKKRA	NAX
	orf26-1	LLGFG	XTGIV	LAVLI	FLLK	DKKRA	NAX
25		490	500				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 (SEQ ID NO: 690) shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) (SEQ ID NO: 696) from *N. gonorrhoeae*:

30	orf26.pep	MQLIDYSHSFFSVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV	60
	orf26ng	MQLIDYSHSFFSVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV	60
	orf26.pep	VGLAWSDXDWSLGGPKILVFXILLGIFTSLTTYSGSN	97
35	orf26ng	VGLAWADGDWSLGGPKILVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		//	
	orf26.pep	TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
	orf26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLAFLCTFGTIKTADYPKA	326
40	orf26.pep	VWQGA	386
	orf26ng	VWQGA	386

```

orf26.pep    ATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR  446
              |||||
orf26ng      ATGTSWGTFGIMLPIAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR  446

orf26.pep    CNHIDHVTSQLPYALTVAAAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK  502
5            |||||
orf26ng      CNHIDHVTSQLPYALTVAAAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKKRADV  506

```

The complete length ORF26ng nucleotide sequence (SEQ ID NO: 695) is:

```

10      1  ATGCAGCTGA  TTGACTATTC  ACATTCATTT  TTCTCGGTTG  TGCCACCCTT
      51  TTTGGCACTG  GCACTTGCCG  TCATTACCCG  CCGCGTACTG  CTGTCTTTAG
     101  GCATCGGTAT  TTTGGTCGGC  GTTGCTTTT  TGGTCGGCGG  CAACCCCGTC
     151  GACGCTCTGA  CACACCTGAA  AGACATGGTC  GTCGCTTGG  CTGGGCAGA
     201  CGGCGATTGG  TCGCTGGGCA  AACCAAAAT  CTTGGTTTTC  CTGATACTTT
     251  TGGGCATTTT  CACTTCACTG  CTGACCTACT  CCGGCAGCAA  TCAGGCGTTT
     15  301  GCCGACTGGG  CAAAACGGCA  CATTAAAAAC  CGGTGCGGCG  CGAAATGCT
     351  GACCGCCTGC  CTCGTGTTTC  TAACCTTTAT  CGACGACTAT  TTCCACAGCC
     401  TCGCCGTCGG  TCGGATTGCC  CGCCCCGTTA  CCGACAAGTT  TAAAGTTTCC
     451  CGCGCCAAAC  TCGCCTACAT  CCTCGACTCC  ACTGCCTCGC  CCATGTGCGT
     501  GCTGATGCCC  GTTTCAGCT  GGGGCGCGTC  GATTATCGCC  ACGCTTGCCG
     20  551  GATTGCTCGT  TACCTACAAA  ATTACCGAAT  ACACGCCGAT  GGGGACGTTT
     601  GTCGCCATGA  GCCTGATGAA  CTATTACGCG  CTGTTTGCCC  TGATTATGGT
     651  ATTCGTCGTC  GCATGGTTCT  CCTTCGACAT  CGGCTCGAtg  gCGCGTTTCG
     701  AACAGGCTGC  GTTGAACGAA  gcccaggacg  aaaccgcccgc  tTCAGACgCT
     751  ACCAAAGGTC  GTGTTTACGC  ATTGATTATT  CCCGTTTTGG  CCTTAATCGC
     25  801  CTCAACGGTT  TCCGCCATGA  TCTACACCGG  CGCGCAGGCA  AGCGAAACCT
     851  TCAGCATTTT  GGGGGCATT  GAAAAATACG  ACGTAAACAC  TTCGCTGGTA
     901  TTCGGCGGCA  CTTGCGGCGT  GCTTGCCGTC  GTCCTCTGCA  CGTTCGGCAC
     951  GATTAAAACC  GCCGATTATC  CAAAGCCGT  GTGGCAGGGT  GCGAAATCCA
    1001  TGTTTCGCGC  AATCGCCATT  TTAATCCTCG  CCTGGCTCAT  CAGTACGGTT
    30  1051  GTCGGCGAAA  TGCACACGGG  CGACTACCTC  TCCACGCTGG  TTGCGGGCAA
    1101  CATCCATCCC  GGCTTCCTGC  CCGTCATCCT  CTTCCTGCTC  GCCAGCGTGA
    1151  TGGCGTTTGC  CACAGGCACA  AGCTGGGGGA  CGTTCGGCAT  TATGCTGCCG
    1201  ATTGCCGCCG  CCATGGCGGT  CAAAGTCGAA  CCCGCGCTGA  TTAtcccGTG
    1251  TATGTCCGCA  GTAATGGCGG  GGGCGGTATG  CGGCGACCAC  TGTTGCCCCA
    35  1301  TCTCCGACAC  GACCATCCTG  TCGTCCACCG  GCGCGCGCTG  CAACCAATC
    1351  GACCACGTTA  CCTCGCAACT  GCCTTATGCC  CTGACGGTTG  CCGCCGCCGC
    1401  CGCATCGGGC  TACCTCGCAT  TGGGTCTGAC  AAAATCCGCG  CTGTTGGGCT
    1451  TTGGCACGAC  CGGTATTGTA  TTGGCGGTGC  TGATTTTCT  GTTGAAGAT
    40  1501  AAAAAACCGC  CCGACGTTTG  A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 696):

```

45      1  MQLIDYSHSF  FSVVPPFLAL  ALAVITRRVL  LSLGIGILVG  VAFLVGGNPV
      51  DGLTHLKDMV  VGLAWADGDW  SLGKPKILVF  LILLGIFTSL  LTYSGSNQAF
     101  ADWAKRHIKN  RCGAKMLTAC  LVFVTFIDY  FHSLAVGAIA  RPVTDKFKVS
     151  RAKLAYILDS  TASPMCVLMP  VSSWGASIIA  TLAGLLVTYK  ITEYTPMGTF
     201  VAMSLMNYA  LFALIMVFVV  AWFSFDIGSM  ARFEQAALNE  AQDETAASDA
     251  TKGRVYALII  PVLALIASTV  SAMIYTGAQA  SETFSILGAF  ENTDVNTSLV
     301  FGGTCGVLA  VLCTFGTIKT  ADYPKAVWQG  AKSMFGAIAI  LILAWLISTV
     351  VGEMHTGDYL  STLVAGNIHP  GFLPVILFLL  ASVMAFATGT  SWGTFGIMLP
     401  IAAAMAVKVE  PALIIPCMSA  VMAGAVCGDH  CSPISDTTIL  SSTGARCNHI
     451  DHVTSQLPYA  LTVAAAAASG  YLALGLTKSA  LLGFGTTGIV  LAVLIFLLKD
     501  KKRADV*

```

ORF26ng (SEQ ID NO: 696) and ORF26-1 (SEQ ID NO: 692) show 98.4% identity in 505 aa overlap:

5	orf26-1.pep	10 20 30 40 50 60	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
	orf26ng	10 20 30 40 50 60	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
10	orf26-1.pep	70 80 90 100 110 120	VGLAWSGDGWSLGGPKILVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRRGAKMLTAC
	orf26ng	70 80 90 100 110 120	VGLAWADGWSLGGPKILVFLILLGIFTSLTTYSGSNQAFADWAKRHIKNRCGAKMLTAC
15	orf26-1.pep	130 140 150 160 170 180	LVFVTFIDDDYFHS LAVGAIARPVTDKFKVSRTKLAYILDSTAAPMCVLMPVSSWGASIIA
	orf26ng	130 140 150 160 170 180	LVFVTFIDDDYFHS LAVGAIARPVTDKFKVSRKLAYILDSTASPMCVLMPVSSWGASIIA
20	orf26-1.pep	190 200 210 220 230 240	TLAGLLVTYKITEYTPMGTFVAMSLMNYIALFALIMVFVVAWFSDIGSMARFEQAALNE
	orf26ng	190 200 210 220 230 240	TLAGLLVTYKITEYTPMGTFVAMSLMNYIALFALIMVFVVAWFSDIGSMARFEQAALNE
25	orf26-1.pep	250 260 270 280 290 300	AHDETAVS DATKGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
	orf26ng	250 260 270 280 290 300	AQDETAAS DATKGRVYALIIPVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
30	orf26-1.pep	310 320 330 340 350 360	FGGTCGVLAVVLCTLGTIKTADYPKAVWQGA KSMFGAIAILILAWLISTVVGEMHTGDYL
	orf26ng	310 320 330 340 350 360	FGGTCGVLAVVLCTFGTIKTADYPKAVWQGA KSMFGAIAILILAWLISTVVGEMHTGDYL
35	orf26-1.pep	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
	orf26ng	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
40	orf26-1.pep	430 440 450 460 470 480	VMAGAVCGDHCSPI SDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
	orf26ng	430 440 450 460 470 480	VMAGAVCGDHCSPI SDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
45	orf26-1.pep	490 500	LLGFGTTGIVLAVLIFLLKDKK RANAX
	orf26ng	490 500	LLGFGTTGIVLAVLIFLLKDKK RADVX

In addition, ORF26 ng (SEQ ID NO: 696) shows significant homology to a hypothetical *H. influenzae* protein(SEQ ID NO: 1156):

```

5      sp|P44263|YF86_HAEIN HYPOTHETICAL PROTEIN HI1586 )gi|1074850|pir||C64037
      .hypothetical
      protein HI1586 - Haemophilus influenzae (strain Rd KW20) )gi|1574427 (U32832) H.
      influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
      Score = 538 bits (1370), Expect = e-152
      Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

10      Query: 1   MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXXAFVGGNPVDGLTHLKDMV 60
      M+LID+S S +S+VP LA+ LA+ TRR L +L V
      Sbjct: 14   MELIDFSSSVWSIVPALLAIILAIATRRVLVSLSAGIIIGSLMLSDWQIGSAFNVLVKNV 73

      Query: 61   VGLAWADGDWSLGPILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120
      V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A
      Sbjct: 74   VSLVYADGEIN-SNMNIVFLLLLLGVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

15      Query: 121 LVFVTFIDDFHSLAVGAIARPVTDFKVSRAKLAYILDSTASPMCVLMPVSSWGASIIA 180
      LVFVTFIDDFHSLAVGAIARPVTD+FKVSRAKLAYILDSTA+PMCV+MPVSSWGA II
      Sbjct: 133 LVFVTFIDDFHSLAVGAIARPVTD+FKVSRAKLAYILDSTAAPMCVMMMPVSSWGAYIIT 192

      Query: 181 TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE 240
      + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL
20      Sbjct: 193 LIGLLATYSITEYTPIGAFVAMSSMNFYAIFSIIMVFFVAYFSFDIASMVRHEKLALKN 252

      Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296
      +D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FENT V
      Sbjct: 253 TEDQLEEETGTGQVRNLILPILVLIATVSMMIYTGAELAADGKVFSVLGTFENTVVG 312

      Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQGAQSMFGXXXXXXXXXXXXSTVVGEM 354
      TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M
25      Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372

      Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSGWTFGIMLPAAAAAMAVKVEPALI 414
      TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLPAAAAAMA P L+
      Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSGWTFGIMLPAAAAAMAANAPELL 432

30      Query: 415 IPCMSAVMAGAVCGDHCSPISDTTILSSTGARC�HIDHVTQXXXXXXXXXXXXXXXXXXXX 474
      +PC+SAVMAGAVCGDHCSP+SDTTILSSTGA+CNHIDHVT+Q
      Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSDTTILSSTGAKCNHIDHVTTQLPYAATVATATSIGYIVV 492

      Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
      S L GF T + L V+IF +K +
35      Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 697):



-507-

```

1  ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
51  CGATGAGCCT GCCAAAATTC TGAATTGGGA TGAAAGCGGC CGATTACTCT
101 CGGAACTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
201 GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCCTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 698; ORF27):

```

1  ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW
51  YEDGSKKSEX VYQDDKLVRK TQWDKDGyli EP*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 699):

```

1  ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
51  GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA
101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAATGGCT
401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAAC GGCGTATGGA AGCAATGGTA
501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCCCTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 700; ORF27-1):

```

1  MKKLSRIVFS TVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV
51  VAGIAHAQDF YPPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
101 KMAGGFSK GK PDGEVWNWYP NGKKSVMMPY KNGLSEGTGY RYYRNGGKES
151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTDW ESRLLSELS
201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF27 (SEQ ID NO: 698) shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) (SEQ ID NO: 702) from strain A of *N. meningitidis*:

```

40  orf27.pep          KQWYADXSIKTEMVMVNDEPAKILTWDESG
                        ||||| : ||||| ||||| ||||| |||||
orf27a  LSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVNDEPAKILTWDESG
          140      150      160      170      180      190

45  orf27.pep          RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDGyliEPX
                        ||||| : ||||| ||||| ||||| ||||| ||||| |||||

```

orf27a RLLSELSIHHRNGVNVLEWYEDGSKKXEAVYQDDKLVRKTQWDXDGYLIEPX  
200 210 220 230 240

The complete length ORF27a nucleotide sequence (SEQ ID NO: 701) is:

```

5      1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
      51 GGCCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAT CAGAACGGGA
     101 AACTGACGGC GACGNTGTCT TCTGCCGCNT ATATCAGGCA ATATAGTGTG
     151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
     201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
    10 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA
     301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
     351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAAATGGTT
     401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC
     451 GAAATCCAGT TTAAACAGAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
    15 501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG
     551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT
     601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG
     651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
    20 701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTGTA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 702):

```

      1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIRQYSV
     51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
    101 KMAGGFSK GK PDGEVWNWYP NGKKS AVMPY KNGLSEGTGX RYYRNGGKES
    25 151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTDW ESGRLLSELS
     201 IHHHRNGVV LEWYEDGSKK XEAVYQDDKL VRKTQWDXDG YLIEP*

```

ORF27a (SEQ ID NO: 702) and ORF27-1 (SEQ ID NO: 700) show 94.7% identity in 245 aa overlap:

```

30      10      20      30      40      50      60
orf27a.pep MKKLSRIVFSTVLLGFSAALPAQXYSVYFNQNGKLTATXSSAAYIRQYSVAEGIAHAQXF
      |||
orf27-1     MKKLSRIVFSTVLLGFSAALPAQYTSVYFNQNGKLTATMSSAAYIRQYSVAVAGIAHAQDF
      10      20      30      40      50      60

35      70      80      90      100     110     120
orf27a.pep XYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSK GKPDGEVWNWYP
      |||
orf27-1     YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSK GKPDGEVWNWYP
      70      80      90      100     110     120

40      130     140     150     160     170     180
orf27a.pep NGKKS AVMPYKNGLSEGTGXRYRNGGKKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN
      |||
orf27-1     NGKKS AVMPYKNGLSEGTGYRYYRNGGKKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
      130     140     150     160     170     180

45      190     200     210     220     230     240
orf27a.pep DEPAKILTWDESGRLLSELSIHHRNGVNVLEWYEDGSKKXEAVYQDDKLVRKTQWDXDG
      |||
orf27-1     DEPAKILTWDESGRLLSELSIRHHQRNGVNVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG
      190     200     210     220     230     240

```

```

orf27a.pep      YLIEPX
                |||||
orf27-1         YLIEPX

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 (SEQ ID NO: 698) shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) (SEQ ID NO: 704) from *N.gonorrhoeae*:

```

10 orf27.pep      KQWYADXS IKTEMVMVNDEPAKILTWDESG 30
    orf27ng      LSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGS IKTEMVMVNDEPAKILTWDESG 193

    orf27.pep      RLLSELSIRHHQRNGVVLWEYEDGSKKSEXVYQDDKLVRKTQWDKDGYLIEP 82
    orf27ng      RLLSELSIRHHKRNGVVLWEYEDGSKKSEAVYQDDKLVRKTQWDKDGYLIEP 245

```

### 15 The complete length ORF27ng nucleotide sequence (SEQ ID NO: 703) is:

```

1  ATGAAGAAAT TATCTCGGAT TGTATTTTCA ATCGTACTGT TGGGTTTTTC
51  GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGGA
101 AACTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
201 151 GCGGCGGGTA TCGCACACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
201 201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AATGGGTCAA
351 351 CTGGTATCCG AACGGTAAAA AATCTGCGGT TATGCCTTAT AAAAATGGCT
401 401 TGAGTGAGGG TACGGGATAC CGTTATTACC GTAACGGCGG CAAGGAAAGC
25 451 GAAATCCAGT TTAAGCAAAA TAAGGCGAAC GCGGTATGGA AGCAATGGTA
501 501 TGCCGATGGA AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTTTC GGAAGTGTCT
601 601 ATCCGCCACC ATAAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
30 651 TTCTAAAAAG AGCGAGGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
701 701 CCAATGGA TAAGGATGGT TATTTAATCG AACCTGA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 704):

```

35 1  MKKLSRIVFS IVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV
51 51  AAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
101 101 KMAGGFSK GK PDGEVWNWYP NGKKS AVMPY KNGLSEGTGY RYYRNGGKES
151 151 EIQFKQNKAN GVKQWYADG SIKTEMVMVN DEPAKILTW ESGRLLSELS
201 201 IRHHRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

```

40 ORF27ng (SEQ ID NO: 704) and ORF27-1 (SEQ ID NO: 700) show 98.8% identity in 245 aa overlap:

```

45 orf27-1.pep      MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF
    orf27ng      MKKLSRIVFSIVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVAAGIAHAQDF

```

-510-

		70	80	90	100	110	120
	orf27-1.pep	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSGKGPGEWVNWYP					
5	orf27ng	YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSGKGPGEWVNWYP					
		70	80	90	100	110	120
	orf27-1.pep	NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN					
10	orf27ng	NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN					
		130	140	150	160	170	180
	orf27-1.pep	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG					
15	orf27ng	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG					
		190	200	210	220	230	240
	orf27-1.pep	YLIEPX					
	orf27ng	YLIEPX					

- 20 Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

25 ORF27-1 (SEQ ID NO: 700) (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 (SEQ ID NO: 700) is a surface-exposed protein and a useful immunogen.

#### Example 84

- 30 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 705):

	1	ATGAAATTTA	CCAAGCACCC	CGTCTGGGCA	ATGGCGTTCC	GCCCATTTTA
	51	TTCGCTGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
	101	GCTACACGGG	AACGCACAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAg
35	151	ATGATTTGGG	GTTATGCCGG	ACTGGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCACT	TGGACGGGGC	AGCCGCCAC	GCGGGGCGGC	GTaTCTGGTC
	251	GGCTTGACTA	TCTTTTGGCT	GGCTGCGCGG	ATTGCCGCCT	TTATCCCGGG
	301	TTGGGGTGCG	TCGGCAAGCG	GCATACTCGG	TACGCTGTTT	TTCTGGTACG
	351	GCGCGGTGTG	CATGGCTTTG	CCCGTTATCC	GTTGCGAGAA	TCAACGCAAC
	401	TATGTTgCCG	TGTTGCGGCT	GTTGCTCTTG	GGCGGCACGC	ATGCGGCGTT
40	451	CCACGTCCAG	CTGCACAACG	GCAACCTAGG	CGGACTCTTG	AGCGGATTGC

501 AGTCGGGCTT GGTGATG

This corresponds to the amino acid sequence (SEQ ID NO: 706; ORF47):

5           1   MKFTKHPVWA   MAFRPFYSLA   ALYGALSVLL   WFGYTGTHX   LSGFYWHAHE  
           51   MIWGYAGLVV   IAFLLTAVAT   WTGQPPTRGG   VLVGLTIFWL   AARIAAFIPG  
          101   WGASASGILG   TLFFWYGAVC   MALPVIRSQN   QRNYVAVFAL   FVLGGTHAAF  
          151   HVQLHNGNLG   GLLSGLQSGS   VM

Further work revealed the complete nucleotide sequence (SEQ ID NO: 707):

10           1   ATGAAATTTA   CCAAGCACCC   CGTCTGGGCA   ATGGCGTTCC   GCCCATTTTA  
           51   TTCGCTGGCG   GCTCTGTACG   GCGCATTGTC   CGTATTGCTG   TGGGGTTTCG  
          101   GCTACACGGG   AACGCACGAG   CTGTCCGGTT   TCTATTGGCA   CGCGCATGAG  
          151   ATGATTTGGG   GTTATGCCGG   ACTGGTCGTC   ATCGCCTTCC   TGCTGACCGC  
          201   CGTCGCCACT   TGGACGGGGC   AGCCGCCAC   GCGGGCGGC   GTTCTGGTCG  
          15   51   GCTTGACTAT   CTTTTGGCTG   GCTGCGCGGA   TTGCCGCCTT   TATCCCGGGT  
          301   TGGGGTGCGT   CGGCAAGCGG   CATACTCGGT   ACGCTGTTTT   TCTGGTACGG  
          351   CGCGGTGTGC   ATGGCTTTGC   CCGTTATCCG   TTCGCAGAA   CAACGCAACT  
          401   ATGTTGCCGT   GTTGCGCGTG   TTCGTCTTGG   GCGGCACGCA   TGCGGCGTTC  
          451   CACGTCCAGC   TGCACAACGG   CAACCTAGGC   GGACTCTTGA   GCGGATTGCA  
          20   501   GTCGGGCTTG   GTGATGGTGT   CGGGTTTAT   CGGTCTGATT   GGTACGCGGA  
          551   TTATTTTCGT   TTTTACGTCC   AAACGCTTGA   ATGTGCCGCA   GATTCCAGT  
          601   CCGAAATGGG   TGGCGCAGGC   TTCGCTGTGG   CTGCCCATGC   TGACTGCCAT  
          651   GCTGATGGCG   CACGGTGTGT   TGGCTTGGCT   GTCTGCCGTT   TTGCCTTTG  
          701   CGGCAGGTGT   GATTTTTACC   GTGCAGGTGT   ACCGCTGGTG   GTATAAACC  
          25   751   GTGTTGAAAG   AGCCGATGCT   GTGGATTCTG   TTTGCCGGCT   ATCTGTTTAC  
          801   CGGATTGGGG   CTGATTGCGG   TCGGCGCGTC   TTATTTCAAA   CCCGCTTTCC  
          851   TCAATCTGGG   TGTGCATCTG   ATCGGGGTCG   GCGGTATCGG   CGTGCTGACT  
          901   TTGGGCATGA   TGGCGCGTAC   CGCGCTTGGT   CATACGGGCA   ATCCGATTTA  
          951   TCCGCCGCC   AAAGCCGTTC   CCGTTGCGTT   TTGGCTGATG   ATGGCGGCAA  
          30   1001   CCGCCGTCCG   TATGGTTGCC   GTATTTTCTT   CCGGCACTGC   CTACACGCAC  
          1051   AGCATCCGCA   CCTCTTCGGT   TTTGTGTTGCA   CTCGCGCTTT   TGGTGTATGC  
          1101   GTGGAAGTAT   ATTCTTGGC   TGATTCTGCC   GCGTTCGGAC   GGCAGCCCC  
          1151   GTTGA

35   This corresponds to the amino acid sequence (SEQ ID NO: 708; ORF47-1):

40           1   MKFTKHPVWA   MAFRPFYSLA   ALYGALSVLL   WFGYTGTHE   LSGFYWHAHE  
           51   MIWGYAGLVV   IAFLLTAVAT   WTGQPPTRGG   VLVGLTIFWL   AARIAAFIPG  
          101   WGASASGILG   TLFFWYGAVC   MALPVIRSQN   QRNYVAVFAL   FVLGGTHAAF  
          151   HVQLHNGNLG   GLLSGLQSGS   VMVSGFIGLI   GTRIISFFTS   KRLNVPQIPS  
          201   PKWVAQASLW   LPMLTAMLMA   HGVLAWSAV   FAFAAGVIFT   VQVYRWYKP  
          251   VLKEPMLWIL   FAGYLFTGLG   LIAVGASYFK   PAFLNLGVHL   IGVGGIGVLT  
          301   LGMARTALG   HTGNPIYPPP   KAVPVAFWLM   MAATAVRMVA   VFSSGTAYTH  
          351   SIRTSSVLFA   LALLVYAWKY   IPWLIRPRSD   GRPG\*

45   Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 (SEQ ID NO: 706) shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) (SEQ ID NO: 710) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf47.pep	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTHXLSGFYWHAHEMIWGYAGLVV					
	orf47a	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYGTHELSGFYWHAHEMIWGYAGLVV					
		10	20	30	40	50	60
10	orf47.pep	70	80	90	100	110	120
		IAFLLTAVATWTGQPPTRGVVLGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC					
	orf47a	IAFLLTAVATWTGQPPTRGVVLGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC					
		70	80	90	100	110	120
15	orf47.pep	130	140	150	160	170	
		MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVM					
	orf47a	MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI					
		130	140	150	160	170	180
20	orf47a	GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVMPWLSAAFAFAAGVIFT					
		190	200	210	220	230	240

The complete length ORF47a nucleotide sequence (SEQ ID NO: 709) is:

	1	ATGAAATTTA	CCAAGCACCC	CGTTTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACGGCG	GCTCTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
25	101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTGGG	GTTATGCCGG	ACTGGTCGTC	ATCGCCTTCC	TGCTGACCGC
	201	CGTCGCCACT	TGGACGGGGC	AGCCGCCAC	GCGGGCGGC	GTTCTGGTCG
	251	GCTTGACTAT	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT
	301	TGGGGTGCGT	CGGCAAGCGG	CATACTCGGT	ACGCTGTTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TTCGCAGAAAT	CAACGCAATT
30	401	ATGTTGCCGT	GTTGCGCGTG	TTCGTCTTGG	GCGGTACGCA	CGCGCGCTTC
	451	CACGTCCAGC	TGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
	501	GTCGGGCTTG	GTGATGGTGT	CGGGTTTTAT	CGGTCTGATT	GGTACGCGGA
	551	TTATTTCTGT	TTTTACGTCC	AAACGGTTGA	ATGTGCCGCA	GATTCCAGT
	601	CCGAAATGGG	TGGCGCAGGC	TTCGCTGTGG	CTGCCCATGC	TGACCGCCAT
35	651	GCTGATGGCG	CACGGCGTGA	TGCCTTGGCT	GTCGGCGGCT	TTCGCGTTTG
	701	CGGCAGGTGT	GATTTTTACC	GTGCAGGTGT	ACCGCTGGTG	GTATAAGCCT
	751	GTGTTGAAAG	AGCCGATGCT	GTGGATTCTG	TTTGCCGGCT	ATCTGTTTAC
	801	CGGATTGGGG	CTGATTGCGG	TGGGCGCGTC	TTATTTCAAA	CCCGCTTTCC
	851	TCAATCTGGG	TGTGCATCTG	ATCGGGGTCG	GCGGTATCGG	CGTGCTGACT
40	901	TTGGGCATGA	TGGCGCGTAC	CGCGCTCGGT	CATACGGGCA	ATCCGATTTA
	951	TCCGCCGCCC	AAAGCCGTTT	CCGTTGCGTT	TTGGCTGATG	ATGGCGGCAA
	1001	CCGCCGTCCG	TATGGTTGCC	GTATTTTCTT	CCGGCACTGC	CTACACGCAC
	1051	AGCATACGCA	CCTCTTCGGT	TTTGTGTTGCA	CTCGCGCTTT	TGGTGTATGC
	1101	GTGGAAGTAT	ATTCCTTGGC	TGATTCTGTC	GCGTTCTGGAC	GGCAGGCCCC
45	1151	GTTGA				

This encodes a protein having amino acid sequence (SEQ ID NO: 710):

	1	MKFTKHPVWA	MAFRPFYSLA	ALYGALSVLL	WGFGYTGTHE	LSGFYWHAHE
50	51	MIWGYAGLVV	IAFLLTAVAT	WTGQPPTRG	VLVGLTIFWL	AARIAAFIPG
	101	WGASASGILG	TLFFWYGAVC	MALPVIRSQN	QRNYVAVFAL	FVLGGTHAAF

5      151   HVQLHNGNLG   GLLSGLQSGL   VMVSGFIGLI   GTRIISFFTS   KRLNVPQIPS  
       201   PKWVAQASLW   LPMLTAMLMA   HGVMPWLSAA   FAFAGVIFT   VQVYRWYKP  
       251   VLKEPMLWIL   FAGYLFTGLG   LIAVGASYFK   PAFLNLGVHL   IGVGGIGVLT  
       301   LGMMARTALG   HTGNPIYPPP   KAVPVAFWLM   MAATAVRMVA   VFSSGTAYTH  
       351   SIRTSSVLFA   LALLVYAWKY   IPWLIRPRSD   GRPG\*

ORF47a (SEQ ID NO: 710) and ORF47-1 (SEQ ID NO: 708) show 99.2% identity in 384 aa overlap:

10	orf47a.pep	10      20      30      40      50      60	MKFTKHPVWAMAFRPFYSLAALYGALS	VLLWGFYGTGTHELSGFYWHAH	EMIWGYAGLVV
	orf47-1		MKFTKHPVWAMAFRPFYSLAALYGALS	VLLWGFYGTGTHELSGFYWHAH	EMIWGYAGLVV
15	orf47a.pep	70      80      90      100      110      120	IAFLLTAVATWTGQPPTRGVLVGLTI	FWLAARIAAFIPGWGASASGILGTL	FFWYGAVC
	orf47-1		IAFLLTAVATWTGQPPTRGVLVGLTI	FWLAARIAAFIPGWGASASGILGTL	FFWYGAVC
20	orf47a.pep	130      140      150      160      170      180	MALPVIRSQNRNYVAVFALFVLGGTH	AAFHVQLHNGNLGGLLSGLQSGL	VMVSGFIGLI
	orf47-1		MALPVIRSQNRNYVAVFALFVLGGTH	AAFHVQLHNGNLGGLLSGLQSGL	VMVSGFIGLI
25	orf47a.pep	190      200      210      220      230      240	GTRIISFFTSKRLNVPQIPSPKWVAQASL	WLPLMTAMLMAHGVMPWLSAAFAFAAGVIFT	
	orf47-1		GTRIISFFTSKRLNVPQIPSPKWVAQASL	WLPLMTAMLMAHGVLAWSAVFAFAAGVIFT	
30	orf47a.pep	250      260      270      280      290      300	VQVYRWYKPVLPKEPMLWILFAGYLFTGL	GLIAVGASYFKPAFLNLGVHLIGVGGIGVLT	
	orf47-1		VQVYRWYKPVLPKEPMLWILFAGYLFTGL	GLIAVGASYFKPAFLNLGVHLIGVGGIGVLT	
35	orf47a.pep	310      320      330      340      350      360	LGMMARTALGHTGNPIYPPP	KAVPVAFWLMMAATAVRMVAVFSSGTAYTHS	SIRTSSVLFA
	orf47-1		LGMMARTALGHTGNPIYPPP	KAVPVAFWLMMAATAVRMVAVFSSGTAYTHS	SIRTSSVLFA
40	orf47a.pep	370      380	LALLVYAWKYIPWLIRPRSDGRPGX		
	orf47-1		LALLVYAWKYIPWLIRPRSDGRPGX		

#### Homology with a predicted ORF from *N.gonorrhoeae*

45      ORF47 (SEQ ID NO: 706) shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) (SEQ ID NO: 712) from *N.gonorrhoeae*:

ORF47 MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV 60  
 |||||  
 ORF47ng MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV 60  
 |||||  
 5 ORF47 IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC 120  
 |||||  
 ORF47ng IAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC 120  
 |||||  
 ORF47 MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM 172  
 |||||  
 10 ORF47ng MALPVIRSQNRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVWGFGLI 180  
 |||||

The ORF47ng nucleotide sequence (SEQ ID NO: 711) is predicted to encode a protein comprising amino acid sequence (SEQ ID NO: 712):

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE  
 51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTAFWL AARIAAFIPG  
 101 WGAAASGILG TLFFWYGAVC MALPVIRSQN RRYVAVFAI FVLGGTHAAAF  
 151 HVQLHNGNLG GLLSGLQSGL VMVWGFGLI GMKIIISFFTS KRLKLPQIPS  
 201 PKWVAHASLW LPMLNAILMA HRVMPWLSAA FPFAAGVIFT VQVYAGGITP  
 251 IEETSCGSVA GICYRLGNSS G

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396 (SEQ ID NO: 1157), accession number e246540):

#### TM segments in ORF47ng

INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence (SEQ ID NO: 713):

1 ATGAAATTTA CCAAACATCC CGTCTGGGCA ATGGCGTTCC GCCCGTTTAA  
 51 TTTACTGGCG GCACTGTACG GCGCATGTGC CGTATTGCTG TGGGGTTTCG  
 101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG  
 151 ATGATTGGG GTTATGCCGG TCTCGTCGTC ATCGCCTTCC TGCTGACCGC  
 201 CGTCGCCACT TGGACGGGAC AGCCGCCAC GAGGGCGGC GTTCTGGTCG  
 251 GCTTGACCGC CTTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGGT  
 301 TGGGGTGCGG CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG  
 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TtcgCAAAC CGGCGCAACT  
 401 ATGtcgCCGT ATTTCGAATA TTTGTGCTGG GCGGTACGCA TGCGgcgTTC  
 451 CACgtccAgc tGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA  
 501 TTCGGGCGCTG GTTATGGTGT CGGGCTTTAT CGGCCTGATT GGGATGAGGA  
 551 TTATTTTCGT TTTTACGTCC AAACGGTTGA ACGTGCCGCA GATTCCCACT  
 601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTACCCATGC TGACCGCCAT  
 651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG  
 701 CGGCGGGCGT GATTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACC  
 751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC



5  
801 CCGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCTGCCTTCC  
851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT  
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATTCGATTTA  
951 TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA  
1001 CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC  
1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC  
1101 GTGGAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG  
1151 GTTGA

10 This encodes a protein having amino acid sequence (SEQ ID NO: 714; ORF47ng-1):

15  
1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE  
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRG VLVGLTAFWL AARIAAFIPG  
101 WGAAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF  
151 HVQLHNGNLG GLLSGLQSG LVMVSGFIGLI GMRIISFFTS KRLNVPQIPS  
201 PKWVAQASLW LPMLTAILMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP  
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT  
301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH  
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG\*

20 ORF47ng-1 (SEQ ID NO: 714) and ORF47-1 (SEQ ID NO: 708) show 97.4% identity in 384 aa overlap:

25  
orf47-1.pep 10 20 30 40 50 60  
MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV  
orf47ng-1 10 20 30 40 50 60  
MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV  
30  
orf47-1.pep 70 80 90 100 110 120  
IAFLLTAVATWTGQPPTRGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC  
orf47ng-1 70 80 90 100 110 120  
IAFLLTAVATWTGQPPTRGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC  
35  
orf47-1.pep 130 140 150 160 170 180  
MALPVIRSONRRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGLVMSGFIGLI  
orf47ng-1 130 140 150 160 170 180  
MALPVIRSONRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLSGLQSGLVMSGFIGLI  
40  
orf47-1.pep 190 200 210 220 230 240  
GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLAHGVLAWSAVFAFAAGVIFT  
orf47ng-1 190 200 210 220 230 240  
GMRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMPWLSAAFAFAAGVIFT  
45  
orf47-1.pep 250 260 270 280 290 300  
VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT  
orf47ng-1 250 260 270 280 290 300  
VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT  
50  
orf47-1.pep 310 320 330 340 350 360  
LGMMARTALGHTGNSIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA  
orf47ng-1 310 320 330 340 350 360  
LGMMARTALGHTGNSIYPPPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA

	310	320	330	340	350	360
	370	380				
orf47-1.pep	LALLVYAWKYIPWLIRPSDGRPGX					
5 orf47ng-1	LALLVYAWKYIPWLIRPSDGRPGX					
	370	380				

Furthermore, ORF47ng-1 (SEQ ID NO: 714) shows significant homology to an ORF (SEQ ID NO: 1157) from *Pseudomonas stutzeri*:

```

10  gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396
    Score = 155 bits (389), Expect = 5e-37
    Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

    Query: 7  PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59
    P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +
15  Sbjet: 14  PIWRLAFRPFFLAGSLYALLAIPLWVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

    Query: 60  VIAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
    V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF
    Sbjet: 72  VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLLFLVALVW 130

    Query: 120 CMALPVIRSQNRNRYVAVFAIFVLGGTHAFAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179
    MA + + +RNY V + ++ G +V+ + L
20  Sbjet: 131 MMAQMLWAVRQKRNYPIVVVLSMLGADVLIILTGLLQNDALQKQVLAGLWLVAALMAL 190

    Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV----MPWLSAFAFA 234
    IG R+I FFT + L P W+ A L + A+L A GV P L F A
    Sbjet: 191 IGGRVIPFFTQKGLGKVDVAVKPVWVLDVALLVGTGVIALHAFGVAMRPQPLLGLLFV-A 249

    Query: 235 AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLFGLGLIAVGASYF-KPAFXXXXXXXXXXX 293
    GV +++ RW+ K + K +LW L L+ + + +F A
25  Sbjet: 250 IGVGHLRLMRWYDKGIWKVGLLWLSHVAMLWLVVAAFGLALWHFGLLAQSSPSLHALSV 309

    Query: 294 XXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXXSSGTAYTHSIR 353
    M+AR LGHTG + P + AF L F S +
30  Sbjet: 310 GSMSGLILAMIAVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAVPVGGLW 365

    Query: 354 TSSVLFALALLVYAWKYIPWLIRPSDGRPG 384
    ++V + LA +Y W+Y P L+ R DG PG
    Sbjet: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 85

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 715):

```

1  ..ATGCCGTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAay CAGAAGyGGT
51  AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG

```



351 IVGNAFGGVG \*

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 86**

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 719)

```

1  ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
51  TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
201 CATGTTTCGCC GCCGGACGAA TTTGGGGGCA GARArTCCTA rGGTTCArAC
251 CTATTGCGsG CATCATGACG CCGrAACGTT ATGAGCAGGT TCAGGAAAAA
301 TTCGACAAAT ACGGTAACGT GGTCTTATTT GTCGCCCCGT TCCTGCCCCG
15  351 TTTGAGAACG GCCGTATTTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
401 ACTTGCGTTT TATCATTATG GATGGACTGG CCGCA...
```

This corresponds to the amino acid sequence (SEQ ID NO: 720; ORF78):

```

1  MFAPLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMYTNP
20  51  HMFVAVGMLG VLVGDGIMFA AGRIWQXXL XFXPIAXIMT PKRYEQVQEK
101  FDYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAA...
```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 721):

```

1  ATGTTTGCTT TTTTAGAAGC CTTTTTTGTC GAATACGGTT ATGCGGCTGT
25  51  TTTTTTTGTA TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101  TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151  CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
201  CATGTTTCGCC GCCGGACGAA TTTGGGGGCA GAAATCCTA AGGTTCAAAC
251  CTATTGCGCG CATCATGACG CCGAAACGTT ATGAGCAGGT TCAGGAAAAA
30  301  TTCGACAAAT ACGGTAACGT GGTCTTATTT GTCGCCCCGT TCCTGCCCCG
351  TTTGAGAACG GCCGTATTTG TTACAGCCGG TATCAGCCGC AAGGTTTCAT
401  ACTTGCGTTT TATCATTATG GATGGACTGG CCGCACTGAT TTCCGTCCCT
451  ATTTGGATTT ATCTGGGCGA ATACGGTGCG CACAACATCG ATTGGCTGAT
501  GGCGAAAATG CACAGCCTGC AATCGGGTAT TTTTGTTATC TTGGGTATAG
35  551  GTGCGACCGT TGTCGCTTGG ATTTGGTGGA AAAAAAGCCA ACGTATCCAG
601  TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
651  GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA
```

This corresponds to the amino acid sequence (SEQ ID NO: 722; ORF78-1):

```

1  MFAPLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMYTNP
40  51  HMFVAVGMLG VLVGDGIMFA AGRIWQKIL RFPKPIARIMT PKRYEQVQEK
101  FDYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
151  IWIYLGHEYGA HNIDWLMAKM HSLQSGIFVI LGIGATVVAV IWWKKRQRIQ
201  FYRSKLKEKR AQRKAAKAAK KAAQSKQ*
```

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H.influenzae* (accession number P45280) (SEQ ID NO: 1158)

- 5 ORF78 (SEQ ID NO: 720) and the dedA homologue (SEQ ID NO: 1158) show 58% aa identity in 144aa overlap:

```

Orf78: 4  FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
          FL  FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+      N H+M  V M+GV
DedA:  20  FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79

10  Orf78: 62  LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
          L  GD  M+  GRI+G  L  F  PI  I+T  R  V+EKF +YGN VLFVARFLPGLR
DedA:  80  LAGDSCMYWLGRIYGTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

Orf78: 122  VFVTAGISRKVSYLRFIIMDGLAA 145
          +++ +GI+R+VSY+RF+++D  AA
15  DedA:  140 IYMSGITRRVSYVRFVLIDFCAA 163

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 (SEQ ID NO: 720) shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) (SEQ ID NO: 724) from strain A of *N. meningitidis*:

```

20  orf78.pep  MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
          10      20      30      40      50      60
orf78a  MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
          10      20      30      40      50      60

25  orf78.pep  VLVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT
          70      80      90      100     110     120
orf78a  VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
          70      80      90      100     110     120

30  orf78.pep  AVFVTAGISRKVSYLRFIIMDGLAA
          130     140
orf78a  AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEGYAHNIDWLMAKMHSLSQGFIA
          130     140     150     160     170     180

```

- 35 The complete length ORF78a nucleotide sequence (SEQ ID NO: 723) is:

```

1  ATGTTTGCCC TTTTGGAAGC CTTTTTTGTC GAATACGGCT ATGCGGCCGT
51 GTTTTTTCGT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101 TGACCTTGGT AACAGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG

```

-520-

5  
10

```

151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
201 CATGTTTCGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA
301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCCG
351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC
451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT
501 GGCGAAAATG CACAGCCTGC AATCCGGCAT CTTCATCGCA TTGGGCGTGC
551 TGGCGCGCGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 724):

15

```

1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMYTNP
51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP
151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRRHYQ
201 LYRAQLSEKR AKRKA EKA KAAQKQ*

```

20 ORF78a (SEQ ID NO: 724) and ORF78-1 (SEQ ID NO: 722) show 89.0% identity in 227 aa overlap:

25

```

      10      20      30      40      50      60
orf78a.pep MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMYTNP HIMFAVGMLG
      ||:|||||||||||||||||||||||||||||||||||||||||||||||||
orf78-1 MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMYTNP HIMFAVGMLG
      10      20      30      40      50      60

      70      80      90      100     110     120
orf78a.pep VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
      |||||||||||||||||||:||||||||||||| |||||||||||||||||||
orf78-1 VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
      70      80      90      100     110     120

      130     140     150     160     170     180
orf78a.pep AVFVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKM HSLQSGIFIA
      |||||||||||||||:|||||||||:||||||||| |||||||||||||||:
orf78-1 AVFVTAGISRKVSYLRFIIMDGLAALISVPWIYLGEYGAHNIDWLMAKM HSLQSGIFVI
      130     140     150     160     170     180

      190     200     210     220
orf78a.pep LGVLAALAWFWRKRRHYQLYRAQLSEKRAKRAEKAAKKAAQKQX
      ||:|::||:|::|::|::|::|::|::|::|::|::|::|::|::|
orf78-1 LGIGATVVAWIWWKKRQRIQFYRSKLKEKRAQRKAAKAAKKAAQSKQX
      190     200     210     220

```

40

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 (SEQ ID NO: 720) shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) (SEQ ID NO: 726) from *N. gonorrhoeae*:

```

orf78.pep    XXLFXPIAXIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF 137
orf78ng      YPVLFFVARFLPGLRTAVFVTAGISRKVSYLRF 32

orf78.pep    IIMDGLAA 145
              :|||||
orf78ng      LIMDGLAALISVPVWIYLGEGYAHNIDWLMAKMHSLSQSGIFIALGVLAALAWFWWRKRR 92

```

The ORF78ng nucleotide sequence (SEQ ID NO: 725) is predicted to encode a protein comprising amino acid sequence (SEQ ID NO: 726):

```

1  ..YPVLFFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
51 GEGYAHNIDW LMAKMHSLSQ GIFIALGVLA AALAWFWWRK RRHYQLYRAQ
101 LSEKRAKRKA EKAAKKAAQK QQ*

```

Further work revealed the complete gonococcal nucleotide sequence (SEQ ID NO: 727):

```

1  atgtttgccc tttTggaagc CTTTTTGTG GAAtacggCt atgcGGCCGT
51 GTTTTTCGTT TTGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT
201 GATGTTTGCC GCCGACGCA TCTGGGGGCA GAAATCCTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
20  301 TTCGACAAAT ACGGCAACTG GGTCTGTTT GTCGCCCCTT TCCTGCCCGG
351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC
451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT
501 GGCGAAAATG CACAGCCTGC AATCGGGCAT CTTTCATCGCA TTGGGCGTGC
25  551 TGGCGCGGCG GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 728; ORF78ng-1):

```

1  MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51 HIMFAVGMGL VLAGDGVMF AGRIWQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP
151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRRHYQ
201 LYRAQLSEKR AKRKA EKA KAAQKQ*

```

ORF78ng-1 (SEQ ID NO: 728) and ORF78-1 (SEQ ID NO: 722) show 88.1% identity in 227 aa overlap:

```

              10      20      30      40      50      60
orf78-1.pep  MFAPLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMGL
              |||:|||||
orf78ng-1    MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMGL
              10      20      30      40      50      60

              70      80      90      100     110     120
orf78-1.pep  VLVGDGIMFAAGRIWQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
              ||:|||||
orf78ng-1    VLAGDGVMF AAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
              70      80      90      100     110     120

```

5 orf78-1.pep AVFVTTAGISRKVSYLRFIIIMDGLAALISVPWIYLLGEYGAHNIDWLMMAKMHSLQSGIFVI  
 orf78ng-1 AVFVTTAGISRKVSYLRFIIIMDGLAALISVPVWIYLLGEYGAHNIDWLMMAKMHSLQSGIFIA  
 10 orf78-1.pep LGIGATVVAWIWWKKRQRIQFYRSKLKEKRAQRKAACAQAQSKQX  
 orf78ng-1 LGVLAAALAWFWWRKRRHYQLYRAQLSEKRAKRAEKAQAQKQX

Furthermore, orf78ng-1 (SEQ ID NO: 728) shows homology to the dedA protein (SEQ ID NO: 1158) from *H.influenzae*:

15 sp|P45280|YG29\_HAEIN HYPOTHETICAL PROTEIN HI1629 )gi|1073983|pir||D64133 dedA  
 protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)  
 )gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212  
 Score = 223 bits (563), Expect = 7e-58  
 Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)  
 20 Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTTLVTGGVISGM--GYTNPHIMFAVGMLGVL 62  
 L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL  
 Sbjct: 21 LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNNSHMLLVSMIGVL 80  
 Query: 63 AGDGVMFAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122  
 AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +  
 Sbjct: 81 AGDSMWLGRIGYTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140  
 25 Query: 123 FVTAGISRKVSYLRFIIIMDGLAALISVPVWIYLLGEYGAHNIDWLMMAKMHSLQSGIFIALG 182  
 ++ +GI+R+VS+Y+RF+++D AA+ISVP+WIIYLLGE GA N+DWL ++ Q I+I +G  
 Sbjct: 141 YMVSGITRRVSYVRFLIDFCAAIISVPWIYLLGELGAKNLDWLHTQIQKGQIVIIYIFIG 200  
 Query: 183 VL 184  
 L  
 30 Sbjct: 201 YL 202

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 87

35 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 729):

40 1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT  
 51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG  
 101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC  
 151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA  
 201 AGTGCATACC CACATCAACG ACAACGCGGT GATGCGGATG CGCGAAGTCG  
 251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC  
 301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA



351 TAAAATTCCT GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC  
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...

This corresponds to the amino acid sequence (SEQ ID NO: 730; ORF79):

5 1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA  
51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELPKG  
101 SYHVMFMGLK KOLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNH..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 731):

10	1	ATGAAAAAAT	TATTGGCGGC	CGTGATGATG	GCAGGTTTGG	CAGGCGCGGT
	51	TTCCGCCGCC	GGAGTCCACG	TTGAGGACGG	CTGGGCGCGC	ACCACCGTCG
	101	AAGGTATGAA	AATAGGCGGC	GCGTTCATGA	AAATCCACAA	CGACGAAGCC
	151	AAACAAGACT	TTTTGTCTCG	CGGAAGCAGC	CCCGTTGCCG	ACCGCGTCGA
	201	AGTGCATACC	CACATCAACG	ACAACGCGGT	GATGCGGATG	CGCGAAGTCG
15	251	AAGGCGGCGT	GCCTTTGGAA	GCGAAATCCG	TTACCGAACT	CAAACCCGGC
	301	AGCTATCATG	TGATGTTTAT	GGGTTTGA	AAACAATTAA	AAGAGGGCGA
	351	TAAATTTCCC	GTTACCCTGA	AATTTAAAAA	CGCCAAAGCG	CAAACCGTCC
	401	AACTGGAAGT	CAAAATCGCG	CCGATGCCGG	CAATGAACCA	CGTTCATCAC
	451	CACGGCGAAG	CGCATCAGCA	CTAA		

This corresponds to the amino acid sequence (SEQ ID NO: 732; ORF79-1):

25

1	<u>MKKLLA</u> <u>VMM</u>	<u>AGLAG</u> <u>VSAA</u>	<u>GVH</u> <u>VED</u> <u>GWAR</u>	<u>TTVE</u> <u>GKIG</u>	<u>AFMK</u> <u>IHNDEA</u>
51	<u>KQDF</u> <u>LLGSS</u>	<u>PVADR</u> <u>VEVHT</u>	<u>HIND</u> <u>NGVM</u>	<u>REVE</u> <u>GGVPLE</u>	<u>AKSV</u> <u>TELKPG</u>
101	<u>SYHV</u> <u>MFMGLK</u>	<u>KQLK</u> <u>EGDKIP</u>	<u>VTLK</u> <u>FKNAKA</u>	<u>QTVQ</u> <u>LEV KIA</u>	<u>PMPAM</u> <u>NHGH</u>
151	<u>HGEA</u> <u>HQH*</u>				

Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF79 (SEQ ID NO: 730) shows 94.6% identity over a 147aa overlap with an ORF (ORF79a)  
 (SEQ ID NO: 734) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf79.pep	<u>MKKLLAAVMMAGLAGAVSAAGVHVEDGWA</u> RTTVEGMKIGGAFMKIHNDEAKQDFLLGGSS					
35			:		:		
	orf79a	<u>MKXLLAAVMMAGLAGAVSAAGIHVEDGWA</u> RTTVEGMKMGGAFMKIHNDEAKQDFLLGGSS					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf79.pep	PVADRVEVH <del>THINDNGVMRMREVEGGVPLEAKSVTELKPGSYHV</del> MF <del>MGLKKQLKEGD</del> KIP					
40							
	orf79a	PVADRVEVH <del>THINDNGVMRMREVEGGVPLEAKSVTELKPGSYHV</del> MF <del>GXXKKQLKXGD</del> KIP					
		70	80	90	100	110	120
		130	140				
	orf79.pep	VT <del>LKFKN</del> AQAQT <del>VQLEV</del> KIAPMPAMNH					

orf79a  
 VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHHGGEAHQH  
 130 140 150

5 The complete length ORF79a nucleotide sequence (SEQ ID NO: 733) is:

1 ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT  
 51 TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG  
 101 AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC  
 151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA  
 10 201 AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG  
 251 AAGCGGCGGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC  
 301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA  
 351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCA CAAACCGTCC  
 401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC  
 15 451 CACGGCGAAG CGCATCAGCA CTAA

This encodes a protein having amino acid sequence (SEQ ID NO: 734):

1 MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TTVEGMKMG AFMKIHND  
 51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG  
 20 101 SYHVMFMGXK QLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGH  
 151 HGEAHQH\*

ORF79a (SEQ ID NO: 734) and ORF79-1 (SEQ ID NO: 732) show 94.9% identity in 157 aa overlap:

25 orf79a.pep 10 20 30 40 50 60  
 MKXLLAAVMMAGLAGAVSAAGIHVEDGWAR TTVEGMKMGAFMKIHND  
 orf79-1 MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHND  
 10 20 30 40 50 60  
 30 orf79a.pep 70 80 90 100 110 120  
 PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKQ  
 orf79-1 PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQ  
 70 80 90 100 110 120  
 35 orf79a.pep 130 140 150  
 VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHHGGEAHQH  
 orf79-1 VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGGEAHQH  
 130 140 150

40 Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 (SEQ ID NO: 730) shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) (SEQ ID NO: 736) from *N.gonorrhoeae*:

-525-

```

orf79.pep    FMKIHNDKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
orf79ng      INDNGVMRMREVKGGVPLEAKSVTELKPGS 30

orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
orf79ng      YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH 86

```

An ORF79ng nucleotide sequence (SEQ ID NO: 735) was predicted to encode a protein comprising amino acid sequence (SEQ ID NO: 736):

```

10      1  ..INDNGVMRM EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
      51  TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence (SEQ ID NO: 737):

```

15      1  ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
      51  TTccgccgCc GGAgTccAtG TCGAggACGG CTGGGCGCGc accaCTGtcg
      101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
      151 atacaaGACt ttgtgcTCgg CGGaaecatg cccgttgccg accgcGTCTGA
      201 AGTGCAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
      251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
20      301 AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAAGTGA AAGAGGGCGA
      351 CAAGATTCCC GTTACCCTGA AATTTAAGAA CGCCAAAGCG CAAACCGTCC
      401 AACTGGAAGT CAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC
      451 CACGGCGAAG CGCATCAGCA CTAA

```

25 This corresponds to the amino acid sequence (SEQ ID NO: 738; ORF79ng-1):

```

30      1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMGG AFMKIHNDKQDFLLGGSS
      51  IQDFVLGGSM PVADRVEVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
      101 SYHVMFMGLK QLKEGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMNHGHH
      151 HGEAHQH*

```

ORF79ng-1 (SEQ ID NO: 738) and ORF79-1 (SEQ ID NO: 732) show 95.5% identity in 157 aa overlap:

```

35      10      20      30      40      50      60
orf79-1.pep  MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDKQDFLLGGSS
orf79ng-1    MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKMGGAFMKIHNDKQDFVLGGSM

      10      20      30      40      50      60

40      70      80      90      100     110     120
orf79-1.pep  PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
orf79ng-1    PVADRVEVHTHINDNGVMRMREVKGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP

      70      80      90      100     110     120

45      130     140     150
orf79-1.pep  VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHHGEAHQH
orf79ng-1    VTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH

      130     140     150

```

Furthermore, ORF79ng-1 (SEQ ID NO: 738) shows significant homology to a protein (SEQ ID NO: 1159) from *Aquifex aeolicus*:

```

5      gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
      Score = 63.6 bits (152), Expect = 6e-10
      Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

      Query: 24  VEDGWARTTVEGMKMGGAFFMKIHNDIAIQDFVLGGSPVADRVEVHTHINDNGVMRMREV 83
                V+  W      G      M I N+      D+++G      +A RVE+H      + +N V +M
      Sbjct: 27  VKHPWVMEPPPGPNTTMMGMIIIVNEGDEPDYLIGAKTDIAQRVELHKTVIENDVAKMVPQ 86

10     Query: 84  KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEV 137
                +  +  + K  E K  YHVM +GLKK++KEGDK+ V L F+ +      TV+  V
      Sbjct: 87  ER-IEIPPKGKVEFKHHGYHVMIIIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

15 ORF79-1 (SEQ ID NO: 732) (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 (SEQ ID NO: 732) is a  
20 surface-exposed protein, and that it is a useful immunogen.

### Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 739):

```

25      1  ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGC GGCCGCA ATATGTTTTG GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
     301  CGGATTCGGG TTGTGAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA
     351  ATacgTGCTG TCCGACAGCA GCCGTTCGTT TAAACGCCG GTA CTCGTGC
     401  CGTTTCCCA GCCCGGTATT TGGACGATyG CTTTCGTGTC AGGGCAGGTG
     451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTTCCGT
     501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
     551  AGAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGACGA AsCATTGAAA
     601  TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
     701  AA

```

# PATENT

# PATENT

# PATENT

# PATENT

# PATENT

# PATENT

# PATENT

# PATENT

# PATENT

# PATENT

		130	140	150	160	170	180
	orf98.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTPNPTGGYY					
5	orf98a	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
		190	200	210	220	230	
	orf98.pep	IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX					
10	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX					
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence (SEQ ID NO: 743) is:

	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
15	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGCCGCA	ATAATGTTTT	GGGTTTAATA	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
	251	CAAACGTATT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CTGTTGGGG
20	301	CGGATTCCGG	TTGTGAAGTC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAAACCA	GTAATCGTGC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTACTAT	ATTATGGTAA
25	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCC	GAACAACAAT
	701	AA				

This encodes a protein having amino acid sequence (SEQ ID NO: 744):

30	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVSXSL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSGQV
	151	SNVKAALPK	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
35	201	YVISLGMVIP	DDLVPVKTLAG	PMPSEKADLP	EQQ*	

ORF98a (SEQ ID NO: 744) and ORF98-1 (SEQ ID NO: 742) show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
40	orf98a.pep	MTEPAAEGGKAAKALKKYLITGILVWLPIAVTWVVSIVSASDQLVNLLPKQWRPQYVL					
	orf98-1	MTEXAAEGGKAAKALKKYLITGILVWLPIAVTWVVSIVSASDQLVNLLPKQWRPQYVL					
		10	20	30	40	50	60
		70	80	90	100	110	120
45	orf98a.pep	GFNIPGLGVIVIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSXSL					
	orf98-1	GFNIPGLGVIVIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL					
		70	80	90	100	110	120
		130	140	150	160	170	180
50	orf98a.pep	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					

orf98-1 SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY  
 130 140 150 160 170 180

orf98a.pep 190 200 210 220 230  
 IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPSEKADLPEQQX  
 |||||  
 orf98-1 IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPSEKADLPEQQX  
 190 200 210 220 230

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 (SEQ ID NO: 740) shows 95.3% identity over a 233 aa overlap with a predicted ORF  
 10 (ORF98ng) (SEQ ID NO: 746) from *N.gonorrhoeae*:

orf98.pep 10 20 30 40 50 60 60  
 MTVTAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL

orf98ng MTEPAAEGGKAAKALKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL 60

orf98.pep 15 GFNIPGLGVIVAIAVLVFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL 120  
 |||||

orf98ng GFNIPGLGVIVAIAVLVFTGLFAANVLGRQILAAWDSLLXRIPVVKSIYSSVKKVSESSL 120

orf98.pep SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY 180  
 |||||

orf98ng SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY 180

orf98.pep IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDLVPKTLAXPMPSEKADLPEQQ 233  
 |||||

orf98ng IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQ 233

25 The complete length ORF98ng nucleotide sequence (SEQ ID NO: 745) is predicted to encode a  
 protein having amino acid sequence (SEQ ID NO: 746):

1 MTEPAAEGGK AAKALKKYLITGILVWLPIA VTVWVVSIV SASDQLVNLL  
 51 PKQWRPQYVL GFNIPGLGVI VAIIVLVFTG LFAANVLGRQ ILAAWDSLLX  
 101 RIPVVKSIYS SVKKVSESSL SDSSRSFKTP VLVFPFQSGI WTIAFVSGQV  
 151 SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK  
 201 YVISLGMVIP DDLVPKTLAG PMPPEKAELP EQQ\*

Further work revealed the complete nucleotide sequence (SEQ ID NO: 747):

1 ATGACGGAAC CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA  
 35 51 ATATCTGATT ACAGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT  
 101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ACCAGCTTGT CAACCTGCTG  
 151 CCGAAGCAAT GCGGCCGCA ATATGTTTGT GGGTTTAATA TCCCCGGCT  
 201 CGGCGTTATT GTTGCCATTG CCGTATTGTT TGTAACCGGA TTATTGCCG  
 251 CAAACGTGTT GGGCCGGCAG ATTCTTGCCG CGTGGGACAG CCTGTTgggg  
 301 cggaTTCCGG TTGTCAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA  
 351 ATCGCTGCTG TCCGACAGCA GCCGTTCGTT TAAAACGCCG GACTCGTG  
 401 CGTTTCCCA ATCGGTATT TGGACAATCG CATTCGTGTC CGGTCAGGTG  
 451 TCGAATGCCG TTAAGGCCGC ATTGCCGAG GATGGCGATT ATCTTTCCGT  
 501 GTATGTCCCG ACCACGCCCA ACCCGACCGG CGTTTACTAT ATTATGGTAA

551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCGTTGAAA  
 601 TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC  
 651 ATTGGCAGGA CCTATGCCGC CTGAAAAGGC GGAGTTGCCG GAACAACAAT  
 701 AA

This corresponds to the amino acid sequence (SEQ ID NO: 748; ORF98ng-1):

1 MTEPAAEGGK AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL  
 51 PKQWRPQYVL GFNIPGLGVI VAIIVLFVTG LFAANVLGRQ ILAAWDSLLG  
 101 RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVPPQSGI WTIAFVSGQV  
 151 SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK  
 201 YVISLGMVIP DDLPVKTLAG PMPPEKAELP EQQ\*

ORF98ng-1 (SEQ ID NO: 748) and ORF98-1 (SEQ ID NO: 742) show 97.9% identity in 233 aa overlap:

15		10	20	30	40	50	60
	orf98-1.pep	MTE	XAA	E	G	G	K
	orf98ng-1	MTE	XAA	E	G	G	K
20		70	80	90	100	110	120
	orf98-1.pep	GFN	IPG	L	G	V	I
	orf98ng-1	GFN	IPG	L	G	V	I
25		130	140	150	160	170	180
	orf98-1.pep	SDS	RSF	K	T	P	V
	orf98ng-1	SDS	RSF	K	T	P	V
30		190	200	210	220	230	
	orf98-1.pep	IMV	K	S	D	V	R
	orf98ng-1	IMV	K	S	D	V	R

Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 89

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 749):

1 ATgAAAACGG TAGTCTGGAT TGTGCTCCTG TTTGCCGCCG CCGTCGGACT



51 GCGCTGGCT TCGGCATTT ACACGGCGA CGTGATATC GTACTCGGAC  
 101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTAGG TTCGCTGATT  
 151 GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GgGgTACTCA  
 201 ATATCCCCGA AAAGATGCAG CGTTTCGGTT CGGCnCGTAA AGGCCkCAAG  
 251 ssCGsGCTTG CCTTGAACAA GGCGGTTTG GCGTATTTTG AAGGGCGTTT  
 301 TGAAAAGGCG GAACTAGAAG CCTCACGCGT GTTGGTCAAC AAAGtAGGCC  
 351 GaGAGACAAC CGGACTTTGG CATTGATGCT GrGCGCGCAC GCCGCCGGAC  
 401 AGATGGAAAA CATCGAssTG CGCGACCGTT ATCTTGCGGA AATCGCCAAA  
 451 CTGCCGGA AAAACAGCAGCT TTCCCGTTAT CTTTGTGTGG CGGAATCGGC  
 10 501 GTTGAACCGG CGCGATTACG AAGCGCGGA AGCCAATCTT CATGCGGCGG  
 551 CGAAGATGAA TGCCAACCTT ACGCGCCTCG TGCGTCTGCA .ATTCGTTAC  
 601 GCTTTCGACA GGGGCGACGC GTTGCAGGTT CTGGCAAAA CCGAAAACT  
 651 TTCCAAGGCG GGCGCGTTGG GCAAATCGGA AATGGAACGG TATCAAAATT  
 701 GGGCATATCC GTCGCCAGCT GGCGGATGCT GCCGATGCCG CCGCTTTGAA  
 15 751 AACCTGCCTG AAGCGGATTC CCGACAGCCT CAAAAACGGG GAATTGAGCG  
 801 TATCGTTGC GGAAAAGTAC GAACGTTTGG GACTGTATGC CGATCGGGTC  
 851 AAATGGGTCA AACAGCATTA TCCGCAsAAC CGCCGCCCG AGCTTTTGGA  
 901 AGCCTTTGTC GAAAGCGTGC GCTTTTGGG CGAGCGCGAA CAGCAGAAAG  
 951 CCATCGATT TGCCGATGCT TGGCTGAAAG AACAGCCCGA TAACGCGCTT  
 20 1001 CTGCTGATGT ATCTCGGTCG GCTCGCCTTC GGCCGCAAA TTTGGGGCAA  
 1051 GGCAAAAGGC TACCTTGAAG CGAGCATTCG ATTAAGCCG AGTATTTCCG  
 1101 CGCGTTTGGT TCTAACAAAG GTTTTCGACG AAATCGGAGA ACCGCAGAAG  
 1151 GCGGAGGCGC AC...

25 This corresponds to the amino acid sequence (SEQ ID NO: 750; ORF100):

1 MKTVVWIVVL FAAAVGLALA SGIYTDVYI VLGQTMLRIN LHAFVLGSLI  
 51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GXKXXLALNK AGLAYFEGRF  
 101 EKAELEASRV LVNVKVRDNR TLALMLXAHA AGQMENIXXR DRYLAEBIAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAMNANLT RLVRLXIRYA  
 30 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSAE KYERLGLYAD AVKWKQHYP XNRRPELLEA  
 301 FVESVRFLGE REQQKAIDFA DAWLKEQPDN ALLLMYLGR LAFGRKLWGKA  
 351 KGYLEASIAL KPSISARLVL TKVFDEIGEP QKAEAH...

35 Further work revealed the complete nucleotide sequence (SEQ ID NO: 751):

1 ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CCGTCGGACT  
 51 GCGCTGGCT TCGGCATTT ACACGGCGA CGTGATATC GTACTCGGAC  
 101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTAGG TTCGCTGATT  
 151 GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GCGTACTCAA  
 201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAA GGCCGCAAGG  
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT  
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGCCGG  
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA  
 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAACTG  
 45 451 CCGGAAAAAC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT  
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA  
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT  
 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAACTTTC  
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG  
 50 701 CATACCGCCG CCAGCTGGCG GATGCTGCCG ATGCCGCCG TTTGAAAACC  
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGGAAT TGAGCGTATC  
 801 GGTTCGGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT  
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC  
 901 TTTGTCGAAA GCGTCGCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT  
 55 951 CGATTTTGCC GATGCTTGCC TGAAAGAACA GCCCGATAAC GCGCTTCTGC  
 1001 TGATGTATCT CGGTCGGCTC GCCTACGGCC GCAAACCTTG GGGCAAGGCA  
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG

1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGGCGG  
 1151 AGGCGCAGCG CAACTTGGTT TTGGAAGCCG TCTCCGATGA CGAACGTCAC  
 1201 GCAGCGTTAG AGCAGCATAG CTGA

5 This corresponds to the amino acid sequence (SEQ ID NO: 752; ORF100-1):

1 MKTVVWIVVL FAAAVGLALA SGIYTGDVYI VLGQTMLRIN LHAFVLGSLI  
 51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF  
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNANLT RLVRQLRYA  
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYP HNRPELLEA  
 301 FVESVRLGE REQQAIDFA DAWLKEQPDN ALLMYLGRL AYGRKLWGKA  
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRNLV LEAVSDDERH  
 401 AALEQHS\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 (SEQ ID NO: 750) shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) (SEQ ID NO: 754) from strain A of *N. meningitidis*:

20		10	20	30	40	50	60
	orf100.pep	MKTVVWIVVLFAAAVGLALASGIYTG	VDVYIVLGQTMLRINLHAFVLGSLI	AVVVWYFLFK			
	orf100a	MKTVVWIVVLFAAAXGLALASGIXTG	VDVYIVLGQTMLRINLHAFVLGSLI	AVVVWYFLFK			
		10	20	30	40	50	60
25		70	80	90	100	110	120
	orf100.pep	FIIGVLNIPEKMQRFGSARKGKXXLALNK	AGLAYFEGRF	EKAELEASRV	LVNKEAGDNR		
	orf100a	FIIGVLNXPEKMQRFGSARKGRKAALALNK	AGLAYFEGRF	EKAELEASRV	LVNKEAGDNR		
		70	80	90	100	110	120
30		130	140	150	160	170	180
	orf100.pep	TLALMLXAHAAAGQMENIXXRDRYLAEIAKL	PEKQQLSRYLLAESALNRRDYEAAEANLH				
	orf100a	TLALMLGAHAAGQMENIELRDRYLAEIAKL	PEKQQLSRYLLAESALNRRDYEAAEANLH				
		130	140	150	160	170	180
35		190	200	210	220	230	240
	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVL	AKTEKLSKAGALGKSEMERYQNWAYRRQLA				
	orf100a	AAAKMNANLTRLVRLQLRYAFDRGDALQVL	AKTEKXSAGAXGKSEMERYQNWAYRRQLX				
		190	200	210	220	230	240
40		250	260	270	280	290	300
	orf100.pep	DAADAAALKTCLKRIPDSLKNGELSVSVAE	KYERLGLYADAVKWKQHYPXNRRPELLEA				
	orf100a	DAADAAALKTCLKRIPDSLKNGELSVSVAE	KYERLGLYADAVKWKQHYPHNRPELLEA				
		250	260	270	280	290	300

-533-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGR	LAFGRKLWGKAKGYLEASIAL				
5	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGR	LAYGRKLWGKAKGYLEASIAL				
		310	320	330	340	350	360
		370	380				
	orf100.pep	KPSISARLVLTKVFDEIGEPQKAEAH					
10	orf100a	KPSISARLVLAKVFDETGEPOKAEAQRNVLASVAEENRPSAETHX					
		370	380	390	400		

The complete length ORF100a nucleotide sequence (SEQ ID NO: 753) is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CNNTCGGGCT
15	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCTGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTCATCATCG	GCGTACTCAA
	201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
20	301	GAAAAGGCGG	AACTTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
	401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG
	451	CCGGAAAAGC	AGCAGCTTTC	CCGTTATCTT	TTGTTGGCGG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
25	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTCGACAGGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
	651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAAACC
	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
30	801	GGTTGCGGAA	AGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAAct	TTTGGAAGCN
	901	TTTGTGCAAA	GCGTGCGCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCCGATAAT	GCGCTTCTGC
	1001	TGANGTATCT	CGGTCGGCTC	GCCTACGGCC	GCAAACCTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTTGTTCTTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCTT
	1201	TCCGCCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence (SEQ ID NO: 754):

40	1	MKTVVWIVVL	FAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNXPE	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAAEANLH	AAAKMNANLT	RLVRLQLRYA
45	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	HNRRPELLEA
	301	FVESVRFLGE	RDQQKAIDFA	DAWLKEQPDN	ALLLXYLGR	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	AKVFDETGE	QKAEAQRNLV	LASVAEENRP
	401	SAETH*				

50 ORF100a (SEQ ID NO: 754) and ORF100-1 (SEQ ID NO: 752) show 95.1% identity in 406 aa overlap:

		10	20	30	40	50	60
	orf100a.pep	MKT	VVWIVVLFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK				
5	orf100-1	MKT	VVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK				
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf100a.pep	FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNGKEAGDNR					
10	orf100-1	FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLVNKEAGDNR					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH					
15	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf100a.pep	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
20	orf100-1	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf100a.pep	DAADAAALKTKLKRI PDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRPELLEA					
25	orf100-1	DAADAAALKTKLKRI PDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRPELLEA					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf100a.pep	FVESVRFLGERDQQAIDFADAWLKEQPDNALLXLYLGRLAYGRKLWGKAKGYLEASIAL					
30	orf100-1	FVESVRFLGEREQQAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
		370	380	390	400		
	orf100a.pep	KPSISARLVLAKVFDETGEPOKAEQRNLVLASVAEENRPSA-ETHX					
35	orf100-1	KPSISARLVLAKVFDEIGEPQKAEQRNLVLEAVSDDERHAALEQHSX					
		370	380	390	400		

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 (SEQ ID NO: 750) shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) (SEQ ID NO: 756) from *N.gonorrhoeae*:

40	orf100.pep	MKT	VVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
	orf100ng	MKT	VVWIVVLFAAAVGLALASGIYTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
	orf100.pep	FIIGVLNIPEKMQRFGSARKGKXXLALNKAGLAYFEGRFEKAELEASRVLVNKGVRDNR		120
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNGKEAGDNR		120

	orf100.pep	TLALMLXAHAAGQMENIXXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
5	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
	orf100ng	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA	240
	orf100.pep	DAADAAALKTKLKRI PDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPXNRRPELLEA	300
	orf100ng	DAADAAALKTKLKRI PDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRRPELLEA	300
10	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360
	orf100ng	FVESVRFLGEREQQKAIDFADSWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360
	orf100.pep	KPSISARLVLTQVFDEIGEPQKAEAH	386
15	orf100ng	KPSIPARLVLAQVFDETAQSQAQAQRNLVLASVAGENRPSAETR	405

The complete length ORF100ng nucleotide sequence (SEQ ID NO: 755) is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
20	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTTAAA	TTCATCATCG	GCGTACTCAA
	201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCTGA	AGGGCGTTTT
	301	GAAAAGGCGG	AACCTGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGGCCGG
25	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACCGG	GCAGGACAGA
	401	TGGAAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAAACTG
	451	CCGGAAAAAC	AGCAGCTTTC	CCGCTATCTT	CTGCTGGCGG	AATCGGCGTT
	501	AAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCC
30	601	TTCGATCGGG	GCGATGCGTT	GCAGGTTCTG	GCAAAAaccG	AAAAACTTTC
	651	CAAGGCGGGC	GCGTTGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGATGGCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAAACC
	751	TGCCTGAAGC	GGATTCCCAG	CAGCCTCAA	AACGGGGAAT	TGagcGTATC
	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
35	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GCCCCGAGCT	TTTGGAAGCC
	901	TTTGTGAAAA	GCGTGCGCTT	TTTGGGCGAG	CGCGAACAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATTCTTGGC	TGAAAGAACA	GCCCCATAAC	GCGCTTCTGC
	1001	TGATGTATCT	CGGCCGGCTC	GCCTACGGCC	GCAAACCTTG	GGGTAAGGCA
	1051	AAAGGCTACC	TTGAAGCGAG	TATTGCACTG	AAGCCGAGTA	TTCCGGCGCG
40	1101	TTTGGTGTG	GCAAAGGTTT	TTGACGAAAC	CGCACAGTCG	CAAAAAGCCG
	1151	AAGCACAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGGGGA	AAACCGCCCT
	1201	TCCGCCGAAA	CCCGTTGA			

This encodes a protein having amino acid sequence (SEQ ID NO: 756):

45	1	MKTVVWIVVL	FAAAVGLALA	SGIYTGdVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNIPE	NMRRSGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQMA	DAADAAALKT
50	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	HNRRPELLEA
	301	FVESVRFLGE	REQQKAIDFA	DSWLKEQPDN	ALLLMYLGR	AYGRKLWGKA
	351	KGYLEASIAL	KPSIPARLV	AKVFDETAQS	QKAEQRNLV	LASVAGENRP

401 SAETR\*

ORF100ng (SEQ ID NO: 756) and ORF100-1 (SEQ ID NO: 752) show 95.3% identity in 402 aa overlap:

5		10	20	30	40	50	60
	orf100-1.pep	MKT	VV	I	V	L	F
	orf100ng	MKT	VV	I	V	L	F
10		70	80	90	100	110	120
	orf100-1.pep	FI	I	G	V	L	N
	orf100ng	FI	I	G	V	L	N
15		130	140	150	160	170	180
	orf100-1.pep	TL	A	L	M	L	G
	orf100ng	TL	A	L	M	L	G
20		190	200	210	220	230	240
	orf100-1.pep	AA	A	K	M	N	A
	orf100ng	AA	A	K	M	N	A
25		250	260	270	280	290	300
	orf100-1.pep	DA	A	D	A	A	A
	orf100ng	DA	A	D	A	A	A
30		310	320	330	340	350	360
	orf100-1.pep	FV	E	S	V	R	F
	orf100ng	FV	E	S	V	R	F
35		370	380	390	400		
	orf100-1.pep	KP	S	I	S	A	R
	orf100n	KP	S	I	P	A	R

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 757)

```

5      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTTCGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
      101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGTCTGTC GGGCATGGCG
      151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
      201 CCGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGCAGC GGCTGGGTAC
      251 ACGTCAAAC GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
      301 GCGGTGCTGC TCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
10     351 CTGGTACCGC GTGTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
      401 TGTATsTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 758; ORF102):

```

15     1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVK PF*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 759):

```

20     1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTTCGTG
      51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
      101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGTCTGTC GGGCATGGCG
      151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
      201 CCGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGCAGC GGCTGGGTAC
      251 ACGTCAAAC GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
25     301 GCGGTGCTGC TCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
      351 CTGGTACCGC GTGTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
      401 TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 760; ORF102-1):

```

30     1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYL VVK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number  
AE000647) (SEQ ID NO: 1160)

ORF102 (SEQ ID NO: 758) and HP1484 (SEQ ID NO: 1160) show 33% aa identity in 143aa overlap:

```

40     orf102  3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      HP1484  8  FLWVKAFHVIAVISWMAALFYLPRLFVYHAENAHKKEFVG VVQIQEK--KLYSFIASPAM 65

```

5

orf102	63	GAVVFGAAIPFAAG---WWGSGVHVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY	119
		G + + + GW+H KL L ++LLAY YC +R + + R+Y	
HP1484	66	GFTLITGIILMLLIEPTLFPKSGGWLHAKLALVVLLLAYHFYCKKCMRELEKDPTRRNARFY	125
orf102	120	RVFNEIPXXXXXXXXXXXXXFKPF	142
		RVFNE P KPF	
HP1484	126	RVFNEAPTILMILIVILVVVKPF	148

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 (SEQ ID NO: 758) shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) (SEQ ID NO: 762) from strain A of *N. meningitidis*:

10

orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL
orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL

15

orf102.pep	GFGAVVFGAAIPFAAGWWGSGVHVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
orf102a	GFGAVVFGAAIPFAAGWWGSGVHVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR

20

orf102.pep	VFNEIPVLLMVAALYXVVKPF
orf102a	VFNEIPVLLMVAALYLXVVKPF

25

The complete length ORF102a nucleotide sequence (SEQ ID NO: 761) is:

30

1	ATGATGTTTT	CTTGTTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTTCGTG
51	GTTTGCAGGG	CTGTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
101	TTGATGTGCC	GCGCGGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
151	GTGCGGCTGT	ACCGTTTTAT	GTCGCCGTTG	GGCTTCGGCG	CGGTCGTGTT
201	CGGCGCGGCG	ATACCGTTTG	CCGCCGGCTG	GTGGGGCAGC	GGCTGGGTAC
251	ACGTCAAAC	GTGTTTGGGC	TTGATGCTCT	TGGCTTACCA	GTTGTATTGC
301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
351	CTGGTACCGC	GTGTTCAACG	AAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

35

This encodes a protein having amino acid sequence (SEQ ID NO: 762):

40

1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVN	MAMIDVPRGN	PEYVRLSGMA
51	VRLYR	FMSPL	GFGAVVFGAA	IPFAAGWWGS	GWVHVHVKLCLG	LMLLAYQLYC
101	G	VLLRRFQDY	SNAF	SHRWYR	VFNEIPVLLM	VAALYLXVVKPF*

ORF102a (SEQ ID NO: 762) and ORF102-1 (SEQ ID NO: 760) show complete identity in 142 aa overlap:



-539-

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL			
5	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL			
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWSG	WVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR		
10	orf102-1	GFGAVVFGAAIPFAAGWWSG	WVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR		
		70	80	90	100	110	120
		130	140				
	orf102a.pep	VFNEIPVLLMVAALYL	VVFKPFX				
15	orf102-1	VFNEIPVLLMVAALYL	VVFKPFX				
		130	140				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 (SEQ ID NO: 758) shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) (SEQ ID NO: 764) from *N. gonorrhoeae*:

20	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	60
	orf102ng	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDAPRGNPEYVRLSGMAVRLYRFMSPL	60
	orf102.pep	GFGAVVFGAAIPFAAGWWSG	WVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR 120
	orf102ng	GFGAVVFGAAIPFAAGRWSG	WVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR 120
25	orf102.pep	VFNEIPVLLMVAALYXV	VFKPF	142	
	orf102ng	VFNEIPVLLMVAALYL	VVFKPF	142	

The complete length ORF102ng nucleotide sequence (SEQ ID NO: 763) is:

30	1	ATGATGTTTT	CTTGTTCAA	GCTGTTTCA	TTGTTTTTTG	TCATTTCTGTG
	51	GTTTGCAGGG	CTGTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
	101	TTGATGCGCC	GCGCGGCAAT	CCCGAGTATG	TGCGCCTGTC	GGGGATGGCG
	151	GTGCGGTTGT	ACCGTTTTAT	GTCGCCTTTG	GGTTTCGGCG	CGGTCGTGTT
	201	CGGCGCGGCG	ATACCGTTTG	CCGCcggccg	GTGGGGCagc	ggctggGTTT
35	251	ACGTCAAAC	TGTTTTGGGC	TTGATGCTCT	TGGCTTATCA	GTTGTATTGC
	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
	351	CTGGTACCGC	GTGTTCAAcg	aaATCCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

40 This encodes a protein having amino acid sequence (SEQ ID NO: 764):

	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVN	MAMIDAPRGN	PEYVRLSGMA
	51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGRWGS	GWVHV	KLCLGLMLLAYQLY	
	101	<u>GVLLRRFQDY</u>	<u>SNAFSHRWYR</u>	<u>VFNEIPVLLM</u>	<u>VAALYL</u>	<u>VVFKPF</u>	<u>PF*</u>

ORF102ng (SEQ ID NO: 764) and ORF102-1 (SEQ ID NO: 760) show 98.6% identity in 142 aa overlap:

```

5      orf102-1.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
      orf102ng      MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
                        10      20      30      40      50      60

10     orf102-1.pep  GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
      orf102ng      GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
                        70      80      90      100     110     120

15     orf102-1.pep  VFNEIPVLLMVAALYL VVFKPFX
      orf102ng      VFNEIPVLLMVAALYL VVFKPFX
                        130     140

```

In addition, ORF102ng (SEQ ID NO: 764) shows significant homology to a membrane protein (SEQ ID NO: 1160) from *H. pylori*:

```

25     gi|2314656 (AE000647) conserved hypothetical integral membrane protein
      [Helicobacter pylori] Length = 148
      Score = 79.2 bits (192), Expect = 1e-14
      Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

      Query: 3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      F W K FH+ VISW A LFYLP R+ FV A + V++ +LY F++
      Sbjct: 8  FLWVKAFHVI AVISWMAALFYLPRL FVYHAENAHKKEFVG VVQIQEK--KLYSFIAS PAM 65

      Query: 63  GAVVFGAAIP-----FAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFQDYSNAFS 115
      G + + F +G GW+H KL L ++LLAY YC +R + +
      Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121

      Query: 116 HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
      R+YRVFNE P KPF
      Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 91

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 765):

```

40     1  ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGGTCG CGGCGGCAGC
      51  GGT TTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
      101 TTACGGAAAC GGTCAGGCGC GGC // .....

```

5  
10  
15  
20  
25  
30  
35  
40  
45

```
//... ATTTCTGTTTA CGATTTTGTG CGAACC GGAT ACGCCGATTA AGGCGAAGCT
51 CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTTACAACA
101 GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCTGTTT
151 GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
201 GGTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
251 TGAAAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTGGG TGCGGACGGC
301 AAGGCGGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
351 CGAAGTAAAA AGCGGGTTGA AAGAGGGGA CAAAGTGGTC ATCTCCGAAA
401 TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACCGCCCT AGGCGGCCGC
451 CCGCGCCGAT AA
```

This corresponds to the amino acid sequence (SEQ ID NO: 766; ORF85):

15  
20

```
1 MAKMMKWA AV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
51 .....
101 .....
151 .....
201 ..... I SFTILSEPDT
251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARSFV PNPDGKLATG
301 MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTMG
351 RDSMNTVEKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*
```

Further work revealed the further partial nucleotide sequence (SEQ ID NO: 767):

25  
30  
35  
40  
45

```
1 ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
51 ACTCGGGCAA CAGGTTAAAA AGGGCGATT TATTGCGGAA ATCAATTCGA
101 CCTCGCAGAC CAATACGCTC AATACGGAAA AATCCAAGTT GGAAACGTAT
151 CAGGCGAAGC TGGTGTGCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
201 ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG
251 ATTTGGAAG CGCGCAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
301 GAGCTGAAGG CTTTAATCAG ACAGAGCAAA ATTTCCATCA ATACCGCCGA
351 GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
401 TGGCGATTCT CGTGGAAGAG GGGCAGACTG TGAACGCGGC GCAGTCTACG
451 CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAATGCA
501 GATTGCCGAG GGCATATTA CCAAGGTGAA GGCGGGGCAG GATATTTCTG
551 TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
601 GTCGACCCCG GGCTGACCAC GATGTCGTCG GGCGGTTACA ACAGCAGTAC
651 GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTTCG TTTGTGCCGA
701 ATCCGGACGG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
751 ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
801 TCGCGGCGGC AAGGCGTTTG TGCGCGTGT GGGTGC GGCAAGGCGG
851 CGGAACGCGA AATCCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
901 AAAAGCGGGT TGAAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
951 CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CCTAGGCGGC CCGCCGCGCC
1001 GATAA
```

45 This corresponds to the amino acid sequence (SEQ ID NO: 768; ORF85-1):

50

```
1 ..VSVGAQASGQ IKILYVKLGQ QVKKGDLIAE INSTSQNTNL NTEKSKLETY
51 QAKLVSAQIA LGSAEKKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA
101 ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
151 PTVQLANLD MMLNKMQIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
201 VDPGLTTMSS GGYNSSTDTA SNAVYYYARS FVPNPDGKLA TGMTTQNTVE
251 IDGVKNVLI PSLTVKNRGG KAFVRVLGAD GKAAEREIRT GMRDSMNTVE
301 KSGLKEGDKV VISEITAAEQ QESGERALGG PPRR*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 (SEQ ID NO: 766) shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) (SEQ ID NO: 770) from strain A of *N. meningitidis*:

5	orf85.pep	MAKMMKWAAVAATAAAVWGGWS-LKPEPHVLDITETVRRG	10	20	30	40
	orf85a	MAKMMKWAAVAATAAAVWGGWSYLKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS	10	20	30	40
			50	60		
10	orf85.pep	.....ISFTILSEPDTPIKAKLDSVDPGLTTMSSG	80	90	100	
	orf85a	TIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG	210	220	230	240
			250	260		
15	orf85.pep	GYNSSTDASNAYYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK	110	120	130	140
	orf85a	GYNSSTDASNAYYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK	270	280	290	300
			310	320		
20	orf85.pep	AFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVEITAAEQQESGERALGGP	170	180	190	200
	orf85a	AFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVEITAAEQQESGERALGGP	330	340	350	360
			370	380		
25	orf85.pep	PRRX	230			
	orf85a	PRRX	390			

The complete length ORF85a nucleotide sequence (SEQ ID NO: 769) is:

1	ATGGCAAAAA	TGATGAAATG	GGCGGCTGTT	GCGGCGGTCTG	CGGCGGCAGC
51	GGTTTGGGGC	GGATGGTCTT	ATCTGAAGCC	CGAGCCGCAG	GCTGCTTATA
101	TTACGGAAAC	GGTCAGGCGC	GGCGACATCA	GCCGGACGGT	TTCTGCAACA
151	GGGGAGATT	CGCCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCATCGGG
201	GCAGATTAAG	AAACTTTATG	TCAAACCTCG	GCAACAGGTT	AAAAAGGGCG
251	ATTTGATTGC	GGAAATCAAT	TCGACCTCGC	AGACCAATAC	GCTCAATACG
301	GAAAAATCCA	AATTGGAAAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
351	TGCATTGGGC	AGCGCGGAGA	AGAAATATAA	GCGTCAGGCG	GCGTTGTGGA
401	AGGATGATGC	GACCGCTAAA	GAAGATTGGG	AAAGCGCACA	GGATGCGCTT
451	GCCGCCGCCA	AAGCCAATGT	TGCCGAGCTG	AAGGCTCTAA	TCAGACAGAG
501	CAAAATTTCC	ATCAATACCG	CCGAGTCGGA	ATTGGGCTAC	ACGCGCATTA
551	CCGCAACGAT	GGACGGCACG	GTGGTGGCGA	TTCTCGTGGA	AGAGGGGCAG
601	ACTGTGAACG	CGGCGCAGTC	TACGCCGACG	ATTGTCCAAT	TGGCGAATCT
651	GGATATGATG	TTGAACAAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
701	TGAAGGCGGG	GCAGGATATT	TCGTTTACGA	TTTTGTCCGA	ACCGGATACG
751	CCGATTAAAG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
801	GTCGGGCGGC	TACAACAGCA	GTACGGATAC	GGCTTCCAAT	GCGGTCTACT
851	ATTATGCCCG	TTCGTTTGTG	CCGAATCCGG	ACGGCAAAC	CGCCACGGGG
901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGCTGAT

5  
951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAGGGCG TTTGTGCGCG  
1001 TGTTGGGTGC AGACGGCAAG GCGGCGGAAC GCGAAATCCG GACCGGTATG  
1051 AGAGACAGTA TGAATACCGA AGTAAAAAGC GGGTTGAAAG AGGGGGACAA  
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC  
1151 GCGCCCTAGG CGGCCCGCCG CGCCGATAA

This encodes a protein having amino acid sequence (SEQ ID NO: 770):

10  
15  
1 MAKMMKWA AV AAVAAA VWG GWSYLKPEPQ AAYITETVRR GDISRTVSAT  
51 GEISPSNLVS VGAQASGQIK KLYVKLGQQV KKGDLIAEIN STSQTNTLNT  
101 EKSKLETYQA KLVSAQIALG SAEKKYKQRA ALWKDDATAK EDLESAQDAL  
151 AAANKANVAEL KALIRQSKIS INTAESELGY TRITATMDGT VVAILVEEGQ  
201 TVNAAQSTPT IVQLANLDM LNKMQIAEGD ITKVKAGQDI SFTILSEPDT  
251 PIKAKLDSVD PGLTTMSSG YNSSTDASN AVYYYARFV PNPDKLATG  
301 MTTQNTVEID GVKNVLIIPS LTVKNRGGRA FVRVLGADGK AAEREIRTGM  
351 RDSMNTVEKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

ORF85a (SEQ ID NO: 770) and ORF85-1 (SEQ ID NO: 768) show 98.2% identity in 334 aa overlap:

20 orf85a.pep 30 40 50 60 70 80  
PQAAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE  
orf85-1 VSVGAQASGQIKILYVKLGQQVKKGDLIAE  
10 20 30  
25 orf85a.pep 90 100 110 120 130 140  
INSTSQTNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKQRAALWKDDATAKEDLESAQD  
orf85-1 INSTSQTNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKQRAALWKENATSKEDLESAQD  
40 50 60 70 80 90  
30 orf85a.pep 150 160 170 180 190 200  
ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST  
orf85-1 AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST  
100 110 120 130 140 150  
35 orf85a.pep 210 220 230 240 250 260  
PTIVQLANLDMMLNKMQUIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS  
orf85-1 PTIVQLANLDMMLNKMQUIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS  
160 170 180 190 200 210  
40 orf85a.pep 270 280 290 300 310 320  
GGYNSSTDASNNAVYYYARFVFPNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG  
orf85-1 GGYNSSTDASNNAVYYYARFVFPNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG  
220 230 240 250 260 270  
45 orf85a.pep 330 340 350 360 370 380  
RAFVRVLGADGKAAEREIRTGMRDSMNTVEKSGLKEGDKVVI SEITAAEQQESGERALGG  
orf85-1 KAFVRVLGADGKAAEREIRTGMRDSMNTVEKSGLKEGDKVVI SEITAAEQQESGERALGG  
280 290 300 310 320 330

```

orf85a.pep  PPRRX
            |||||
orf85-1     PPRRX

```

- 5 Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a (SEQ ID NO: 770).

Homology with a predicted ORF from *N.gonorrhoeae*

ORF85 (SEQ ID NO: 766) shows a high degree of identity with a predicted ORF (ORF85ng) (SEQ ID NO: 772) from *N.gonorrhoeae*:

```

10  ORF85      1 MAKMMKWAAVAAVAAAAVWGGWS.LKPEPHVLDITETVRRG..... 40
    ORF85ng    1 MAKMMKWAAVAAVAAAAVWGGWSYLKPEPQAAYITEAVRRGDISRTVSAT 50
                                     |||||: |||: |||
15  ORF85      .....ISFTILSEPDT 250
    ORF85ng    201 TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT 250
                                     |||||
20  ORF85      251 PIKAKLDSVDPGLTTMSSGGYNSSTD TASNAVYYYARSFVPNPDGKLATG 300
    ORF85ng    251 PIKAKLDSVDPGLTTMSSGGYNSSTD TASNAVYYYARSFVPNPDGKLATG 300
                                     |||||
25  ORF85      301 MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM 350
    ORF85ng    301 MTTQNTVEIDGVKNVLLIPSLTVKNRGGKAFVRVLGADGKAVEREIRTGM 350
                                     |||||
30  ORF85      152 RDSMNTVEKSGLKEGDKV VISEITAAEQQESGERALGGPPRR 393
    ORF85ng    351 KDSMNTVEKSGLKEGDKV VISEITAAEQQESGERALGGPPRR 393
                                     :|||

```

- 30 The complete length ORF85ng nucleotide sequence (SEQ ID NO: 771) is:

```

1  ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CGGCGGCaac
51 GGT TTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAaac ggTCAGGCGC GCGGATATCA GCCGACGGT TTCCGCGACG
151 GgcgAGATT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
35 201 GCAGATTAAA AAGCTTTATG TCAAAC TCGG GCAACAGGTC AAAAAGGGCG
251 ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
401 AGGATGATGC GACCTCTAAA GAAGATTGG AAAGCGCGCA GGATGCGCTT
40 451 GCCGCCGCCA AAGCCAATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501 CAAAATTTC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATT
551 CCGCGACGAT GGACGGCAGC GTGGTGGCGA TTCCCGTGGA AGAGGGGCAG
601 ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
651 GGATATGATG TTGAACAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
45 701 TGAAGGCGGG GCAGGATATT TCGTTACGA TTTGTCCGA ACCGGATACG
751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851 ATTATGCCC TTCTTTGTG CCGAATCCGG ACGCAAACCT CGCCACGGGG
901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGTTGCT

```

-545-

5  
951 TATTCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACGCG  
1001 TGTTGGGTGC GGACGGCAAG GCAGTGGAAC GCGAAATCCG GACCGGTATG  
1051 AAAGACAGTA TGAATACCGA AGTGAAAAGC GGGTTGAAAG AGGGGGACAA  
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC  
1151 GCGCCCTAGG CGGCCCGCCG CGCCGATAA

This encodes a protein having amino acid sequence (SEQ ID NO: 772):

10  
15  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

ORF85ng (SEQ ID NO: 772) and ORF85-1 (SEQ ID NO: 768) show 96.1% identity in 334 aa overlap:

20  
30 40 50 60 70 80  
orf85ng PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKLYVKLGQQVKKGDLIAE  
orf85-1 VSVGAQASGQIKLYVKLGQQVKKGDLIAE  
10 20 30

25  
90 100 110 120 130 140  
orf85ng INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD  
orf85-1 INSTSQNTLNTTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD  
40 50 60 70 80 90

30  
150 160 170 180 190 200  
orf85ng ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST  
orf85-1 AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST  
100 110 120 130 140 150

35  
210 220 230 240 250 260  
orf85ng PTIVQLANLDMMLNKMQUIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS  
orf85-1 PTIVQLANLDMMLNKMQUIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS  
160 170 180 190 200 210

40  
270 280 290 300 310 320  
orf85ng GGYNSSTDITASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG  
orf85-1 GGYNSSTDITASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG  
220 230 240 250 260 270

45  
330 340 350 360 370 380  
orf85ng KAFVRLGADGKAVEREIRTGMKDSMNTTEVKSGLKEGDKVVISITAAEQQESGERALGG  
orf85-1 KAFVRLGADGKAVEREIRTGMKDSMNTTEVKSGLKEGDKVVISITAAEQQESGERALGG  
280 290 300 310 320 330

```
          390
orf85ng    PPRRX
           |||||
orf85-1    PPRRX
```

In addition, ORF85ng (SEQ ID NO: 772) shows significant homology to an *E.coli* membrane fusion protein (SEQ ID NO: 1161):

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from membrane fusion protein precursor, MTRC\_NEIGO SW: P43505 (412 aa) [Escherichia coli] Length = 380  
Score = 193 bits (485), Expect = 2e-48  
Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

Query: 29 PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE 88  
P Y T VR GD+ ++V ATG++ V VGAQ SGQ+K L V +G +VKK L+  
Sbjct: 41 PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTLVAIGDKVKKDQLLGV 100

Query: 89 INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRAALWKDDATSKEXXXXXXX 148  
I+ N I ++ L +A+ A+ L A Y RQ L + A S++  
Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAEAEKLARVTYSRQRLAQTKAVSQQDLDTAAT 160

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVAIPVEEGQTVNAAQST 208  
I++++ S++TA+++L YTRI A M G V I +GQTV AAQ  
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQQAA 220

Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTMMSS 268  
P I+ LA++ ML K Q++E D+ +K GQ FT+L +P T + ++ V P  
Sbjct: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDPLTRYEGQIKDVLP----- 273

Query: 269 GGYNSSTDASNAVYYYARSFVPNPDKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328  
+ + ++A++YYAR VNP+G L MT Q +++ VKNVL IP + + G  
Sbjct: 274 -----TPEKVNDAlFYARFEVFPNPNGLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVI SE 372  
+V L +G+ ERE+ G ++ + E+ GL+ GD+VVI E  
Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDTDVEIVKGLEAGDEVVIGE 373

Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (SEQ ID NO: 768) (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 (SEQ ID NO: 768) is a surface-exposed protein, and that it is a useful immunogen.



**Example 92**

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 773):

```

1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
5  101  CGGTGTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
151  GGCAAACGTG ATGCGGAAGc CAAATTTCGCC GACgGcAGCG TAACTTACGG
201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAGGCT ATGGATTTGT
251  TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCGGGG
10  301  CTGAAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA
351  GGCGGGTACA GGAAAATACA GCATAGGCGG CGTGGAACC GAAGTCGTCA
401  AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATTT cTTCGCACCG
451  TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC
501  CTATACGCTG AAACCTCAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC
15  551  CGTAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 774; ORF120):

```

1  ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR
51  GKLYAEAKFA DGSVTYGKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
20  101  LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
151  SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP*

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 775):

```

1  ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
25  51  CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
101  ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
151  AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201  TTTCGAGTCC GGCGGTACGG TTGTCCGCAA TACCCTGCAC CCTACCTACT
251  ATAGAGACAT ACGCAGGGGC AAACGTATG CGGAAGCCAA ATTCGCCGAC
30  301  GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
351  CAAGGCTATG GATTTGTTCA CGCTTGCTTG GCAGTTGGCG GCAAATGACG
401  CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
451  GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
501  GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
551  TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
35  601  ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651  CGGCCAGGCA GCCAAACCGT AA

```

This corresponds to the amino acid sequence (SEQ ID NO: 776; ORF120-1):

```

1  MMKTFKNIFS AAILSAAALPC AYAAGLPQSA VLHYSGSYGI PATMTFERSG
40  51  NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
101  GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151  VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
201  TDDGKTYTLK LKSVQINGQA AKP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 (SEQ ID NO: 774) shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) (SEQ ID NO: 778) from strain A of *N. meningitidis*:

```

5      orf120.pep      IPATMTFERSGNAYKIVSTIKVPLYNIRFE
      orf120a      SAAILSAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE
      10      20      30      40      50      60

10     orf120.pep      SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAGESKTEQSPKAMD LFTLAWQL
      orf120a      SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAXXXXXXQSPKAMD LFTLAWQL
      70      80      90      100      110      120

15     orf120.pep      AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP
      orf120a      AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP
      130      140      150      160      170      180

20     orf120.pep      SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
      orf120a      SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
      190      200      210      220

```

The complete length ORF120a nucleotide sequence (SEQ ID NO: 777) is:

```

25      1  ATGATGAAGA  CTTTAA AAAA  TATATTTTCC  GCCGCCATTT  TGTCCGCCCG
      51  CCTGCCGTGC  GCGTATGCGG  CAGGGCTGCC  CNAATCCGCC  GTGCTGCACT
      101  ATTCCGGCAG  CTACGGCATT  CCCGCCACNA  NNANNTNNGN  ACNNNGNGNC
      151  AATGCTTNCA  AAATCGTTTC  GACGATTAAA  GTGCCGCTAT  ACAATATCCG
      201  TTTTCGAGTC  GCGGTACGG  TTGTCGGCAA  TACCCTGCAC  CCTACCTACT
      251  ATAGAGACAT  ACGCAGGGGC  AAACGTGTATG  CGGAAGCCAA  ATTCGCCGAC
      301  GGCAGCGTAA  CCTACGGCAA  AGCGGNNNNN  ANCNNNNNNG  NGCAAAGCCC
      351  CAAGGCTATG  GATTGTGTTCA  CGCTTGCNTG  GCAGTTGGCG  GCAAATGACG
      401  CGAAACTCCC  CCCGGGGCTG  AAAATCACCA  ACGGCAAAAA  ACTTTATTCC
      451  GTCGGCGGTT  TGAATAAGGC  GGGTACAGGA  AAATACAGCA  TAGGCGGCGT
      501  GGAAACCGAA  GTCGTCAAAT  ATCGGGTGCG  GCGCGGCAC  GATGCGGTAA
      551  TGTATTCTT  CGCACCGTCC  CTGAACAATA  TTCCGGCACA  AATCGGCTAT
      601  ACCGACGACG  GCAAAACCTA  TACGCTGAAA  CTCAAATCGG  TGCAGATCAA
      651  CGGCCAGGCA  GCCAAACCGT  AA

```

40 This encodes a protein having amino acid sequence (SEQ ID NO: 778):

```

45      1  MMKTFKNIFS  AAILSAA LPC  AYAAGLPXSA  VLHYSGSYGI  PATXXXXXXX
      51  NAXKIVSTIK  VPLYNIRFES  GGTVVGNTLH  PTYYRDIRRG  KLYAEAKFAD
      101  GSVTYGKAXX  XXXXQSPKAM  DLFTLAWQLA  ANDAKLPPGL  KITNGKKLYS
      151  VGGLNKAGTG  KYSIGGVETE  VVKYRVRRGD  DAVMYFFAPS  LNNIPAQIGY
      201  TDDGKTYTLK  LKSVQINGQA  AKP*

```

ORF120a (SEQ ID NO: 778) and ORF120-1 (SEQ ID NO: 776) show 93.3% identity in 223 aa overlap:

-549-

		10	20	30	40	50	60
	orf120a.pep	MMKTFKNIFSAAILSAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIK					
5	orf120-1	MMKTFKNIFSAAILSAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf120a.pep	VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTY GKAXXXXXXQSPKAM					
10	orf120-1	VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTY GKAGESKTEQSPKAM					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf120a.pep	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD					
15	orf120-1	DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD					
		130	140	150	160	170	180
		190	200	210	220		
	orf120a.pep	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX					
20	orf120-1	DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAPX					
		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 (SEQ ID NO: 774) shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) (SEQ ID NO: 780) from *N.gonorrhoeae*:

25	orf120.pep	IPATMTFERSGNAYKIVSTIKVPLYNIRFE	30
	orf120ng	SAAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIKVPLYNIRFE	69
	orf120.pep	SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTY GKAGESKTEQSPKAMD LFTLAWQL	90
	orf120ng	SGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTY GKAGESKTEQSPKAMD LFTLAWQL	129
30	orf120.pep	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP	150
	orf120ng	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP	189
	orf120.pep	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAP	184
35	orf120ng	SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAP	223

The complete length ORF120ng nucleotide sequence (SEQ ID NO: 779) is:

	1	ATGATGAAGA	CTTTTAAAAA	TATATTTTCC	GCCGCCATTT	TGTCCGCCGC
	51	CCTGCCGTGC	CGGTATGCGG	CAAGGCTACC	CCAATCCGCC	GTGCTGCACT
40	101	ATTCCGGCAG	CTACGGCATT	CCCGCCACGA	TGACATTGTA	ACGCAGCGGC
	151	AATGCTTACA	AAATCGTTTC	GACGATTAAA	GTGCCGCTAT	ACAATATCCG
	201	TTTCGAATCC	GGCGGTACGG	TTGTCGGCAA	TACCTGCAC	CCTGCCTACT
	251	ATAAAGACAT	ACGCAGGGGC	AAACTGTATG	CGGAAGCCAA	ATTCGCCGAC
	301	GGCAGCGTAA	CCTACGGCAA	AGCGGGCGAG	AGCAAAACCG	AGCAAAGCCC
45	351	CAAGGCTATG	GATTTGTTCA	CGCTTGCCTG	GCAGTTGGCG	GCAATGACG

-550-

5  
 401 CGAAACTCCC CCCGGGTCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC  
 451 GTCGGCGGCC TGAATAAGGC GGGTACGGGA AAATACAGCA TaggCGGCGT  
 501 GGAAACCGAA GTCGTCAAAT ATCGGGTGC GCGCGGCGAC GATACGGTAA  
 551 CGTATTTCTT CGCACCCTCC CTGAACAATA TTCCGGCACA AATCGGCTAT  
 601 ACCGACGACG GCAAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA  
 651 CGGACAGGCC GCCAAACCGT AA

This encodes a protein having amino acid sequence (SEQ ID NO: 780):

10  
 1 MMKTFKNIFS AAILSAAALPC AYAARLPQSA VLHYSGSYGI PATMTFERSG  
 51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PAYYKDIRRG KLYAEAKFAD  
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS  
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY  
 201 TDDGKTYTLK LKSVQINGQA AKP\*

15 In comparison with ORF120-1 (SEQ ID NO: 776), ORF120ng (SEQ ID NO: 780) shows 97.8% identity in 223 aa overlap:

		10	20	30	40	50	60
orf120-1.pep		MMKTFKNIFS	AAILSAAALPC	AYAARLPQSA	VLHYSGSYGI	PATMTFERSG	NAYKIVSTIK
20	orf120ng	MMKTFKNIFS	AAILSAAALPC	AYAARLPQSA	VLHYSGSYGI	PATMTFERSG	NAYKIVSTIK
		10	20	30	40	50	60
		70	80	90	100	110	120
orf120-1.pep		VPLYNIRFES	GGTVVGNTLH	PYYKDIRRG	KLYAEAKFAD	GSVTYGKAGE	SKTEQSPKAM
25	orf120ng	VPLYNIRFES	GGTVVGNTLH	PYYKDIRRG	KLYAEAKFAD	GSVTYGKAGE	SKTEQSPKAM
		70	80	90	100	110	120
		130	140	150	160	170	180
orf120-1.pep		DLFTLAWQLA	ANDAKLPPGL	KITNGKKLYS	VGGLNKAGTG	KYSIGGVETE	VVKYRVRRGD
30	orf120ng	DLFTLAWQLA	ANDAKLPPGL	KITNGKKLYS	VGGLNKAGTG	KYSIGGVETE	VVKYRVRRGD
		130	140	150	160	170	180
		190	200	210	220		
orf120-1.pep		DAVMYFFAPS	LNNIPAQIGY	TDDGKTYTLK	LKSVQINGQA	AKPX	
35	orf120ng	DTVTYFFAPS	LNNIPAQIGY	TDDGKTYTLK	LKSVQINGQA	AKPX	
		190	200	210	220		

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
 40 useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 93

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 781):

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC

51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTTCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT  
 201 GATGGTGTTC TCCTTGATTT TGTGTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence (SEQ ID NO: 782; ORF121):

1 MYRRKGRGIK PWMGAGXAFALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNI..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 783):

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGATGG GTGCCGGTGC  
 51 GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTTCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT  
 201 GATGGTGTTC TCCTTGATTT TGTGTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC  
 501 CTTGCTGCTT TACTATTTCC TGCTGGATTG GCAGCGGTGG TCGTGCGGCA  
 551 TTGCCAAACT GGTTCGAGG CGTTTGGCG GTGCTTATAC GCGATTACA  
 601 GGCAATTGGA ACGAGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT  
 651 AATGCTGATT ATGGGCTTGG TTTACGTTT GGGATTGGTG CTGGTCGGGC  
 701 TGGATTCGGG GTTTGCCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTC  
 751 CCTTATCTCG GGGCTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCTT  
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG  
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACGCCGAA AATCGTGGGA  
 901 GACCGTATCG GGTGTCGCC GTTTTGGGT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTGG TCGGAATGTT GGCGGGATTG CTTTGGCCG  
 1001 CCGTAACCTT GGTCTGCTT CCGGAGGGCG TGCAGAAATA TTTTGCCGGC  
 1051 AGTTTTTACC GGGCAGGTA G

This corresponds to the amino acid sequence (SEQ ID NO: 784; ORF121-1):

1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNIVSSI GNLLLPPLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT  
 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVFV  
 251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG  
 351 SFYRGR\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF121 (SEQ ID NO: 782) shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) (SEQ ID NO: 786) from strain A of *N. meningitidis*:

5	orf121.pep	10 20 30 40 50 60	MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR
	orf121a	10 20 30 40 50 60	MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR
10	orf121.pep	70 80 90 100 110 120	ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
	orf121a	70 80 90 100 110 120	ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
15	orf121.pep	130 140 150	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
	orf121a	130 140 150	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
20	orf121a	130 140 150 160 170 180	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
	orf121a	190 200 210 220 230 240	SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI

The complete length ORF121a nucleotide sequence (SEQ ID NO: 785) is:

25	1	ATGTATCGGA	GGAAAGGGCG	GGGCATCAAG	CCGTGGATGG	ATGCCGGTGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
30	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTGT
35	201	GATGGTGTTC	TCCTTGATTT	TGTTGTGGC	ATTATTGTTG	ATTATTGTCC
	251	CTATGCTGGT	CGGGCAGTTC	AACAATTTGG	CATCGCGCCT	GCCCCAATTA
40	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	CTTCAGGCGC
45	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
	451	AGGCAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCTTCC
50	501	CTTGCTGCTT	TACTATTTCC	TGCTGGATTG	GCAGCGGTGG	TCGTGCGGCA
	551	TTGCCAAACT	GGTCCGAGG	CGTTTTGCCG	GTGCTTATAC	GCGCATTACA
55	601	GGCAATTTGA	ACGCGTATT	GGGCGAATTT	TTGCGCGGGC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGTTTGG	TTTACGGCTT	GGGGTTGGTG	CTGGTCGGGC
60	701	TGGATTCGGG	GTTTGCAATC	GGTATGGTTG	CCGGTATTTT	GGTTTTTGTT
	751	CCCTATTTGG	GCGCGTTTAC	AGGACTGCTG	CTGGCAACCG	TCGCCGCCTT
65	801	GCTCCAGTTC	GGTTCGTGGA	ACGGCATCTT	GGCTGTTTGG	GCGGTTTTTG
	851	CCGTAGGACA	GTTTCTCGAA	AGTTTTTTCA	TTACGCCGAA	AATCGTGGGA
70	901	GACCGTATCG	GCCTGTCGCC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGGCAGCTG	ATGGGCTTTG	TGCGAATGTT	GGCCGGATTG	CCTTTGGCCG
75	1001	CCGTAACTTT	GGTCTTGCTT	CGCGAGGGCG	TGCAGAAATA	TTTTGCCGGC
	1051	AGTTTTTACC	GGGGCAGGTA	G		

45 This encodes a protein having amino acid sequence (SEQ ID NO: 786):

50	1	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV
	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL
55	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	LQAHTGELSN	ALKAWFPVLM
	151	RQGGNIVSSI	GNLLLLPLLL	YYFLLDWQRW	SCGIAKLVPR	RFAGAYTRIT
60	201	GNLNEVLGEF	LRGQLLVMLI	MGLVYGLGLV	LVGLDSGFAI	GMVAGILVVF

251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG  
 351 SFYRGR\*

5 ORF121a (SEQ ID NO: 786) and ORF121-1 (SEQ ID NO: 784) show 99.2% identity in 356 aa overlap:

		10	20	30	40	50	60
	orf121a.pep	MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR					
10	orf121-1	MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR					
		10	20	30	40	50	60
	orf121a.pep	ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV					
15	orf121-1	ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV					
		70	80	90	100	110	120
	orf121a.pep	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW					
20	orf121-1	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW					
		130	140	150	160	170	180
	orf121a.pep	SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI					
25	orf121-1	SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI					
		190	200	210	220	230	240
	orf121a.pep	GMVAGILVFVPYLGAFITGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG					
30	orf121-1	GMLAGILVFVPYLGAFITGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG					
		250	260	270	280	290	300
	orf121a.pep	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX					
35	orf121-1	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX					
		310	320	330	340	350	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 (SEQ ID NO: 782) shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) (SEQ ID NO: 788) from *N.gonorrhoeae*:

40	orf121.pep	MYRRKGRGIKPWMGAGXAAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR	60
	orf121ng	MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYVLDPLVEWLQKKGLNR	60

```

orf121.pep  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf121ng    ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV  120

orf121.pep  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI  156
             |||||||||:|||||||||||||||||||:|||||
5 orf121ng  EIDQASIIAWFQAHTGELSNAKAWFPVLMKQGGNIVSTIGNLLLPPLLLYYFLLDWHRW  180

```

An ORF121ng nucleotide sequence (SEQ ID NO: 787) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 788):

```

10      1  MYRRKGRGIK  PWMGAGAAFA  ALVWL VYALG  DTLTPFAVAA  VLAYVLDPLV
      51  EWLQKGLNR  ASASMSVMVF  SLILL LALLL  IIVPMLVGQF  NNLASRLPQL
     101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  FQAHTGELSN  ALKAWFPVLM
     151  KQGGNIVSTI  GNLLLPPLLL  YYFLLDWHRW  SCGIPKLVR  RFAGAYTRIT
     201  GNLNKVGWKF  LRGQLLGETE  RGA VVCRVGR  ECWEGGGARS  RPSDDGWPRW
15     251  GGG*

```

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 789):

```

      1  ATGTATCGGA  GAAAAGGACG  GGGCATCAAG  CCGTGGATGG  GTGCCGGCGC
     51  GGCCTTTGCC  GCCTTGGTCT  GGCTGGTTTA  CGCGCTCGGC  GATACTTTGA
20    101  CTCCTTTGTC  GGTTGCGGCG  GTGCTGGCGT  ATGTGTTGGA  CCCTTTGGTC
     151  GAATGGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGTCTGT
     201  GATGGTGTTT  TCCTTGATTT  TGTGTGTGGC  ATTATTGTTG  ATTATTGTCC
     251  CTATGCTGGT  CGGGCAGTTC  AATAATTTGG  CATCTCGCCT  GCCCAATTA
     301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAAA  ATACAATCGG
25    351  CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  TTTCAGGCGC
     401  ATACGGGCGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTGTATG
     451  AAACAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCCGCC
     501  CTTGCTGCTT  TACTATTTCC  TGCTGGATTG  GCAGCGGTGG  TCGTGCGGCA
     551  TCGCCAAACT  GGTTCGAGG  CGTTTGGCCG  GTGCTTATAC  GCGCATTACG
30    601  GGTAATTGTA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGTC  AGCTTCTGGT
     651  GATGCTGATT  ATGGGCTTGG  TTTACGGTTT  GGGATTGATG  CTAGTCGGAC
     701  TGGATTCGGG  ATTTGCCATC  GGTATGGTTG  CCGGTATTTT  GGTGTTTGTC
     751  CCCTATTGCG  GTGCGTTTAC  GGGATTGCTG  CTTGCCACTG  TTGCAGCCTT
     801  GCTCCAGTTC  GGTTCGTGGA  ACGGAATCTT  GGCTGTTTGG  GCGGTTTTTG
35    851  CCGTCGGTCA  GTTCTCGAA  AGTTTTTTCA  TTACGCCGAA  AATTGTAGGA
     901  GACCGTATCG  GCCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
     951  CGGAGAGCTG  ATGGGCTTTG  TCGGAATGTT  GGCCGGATTG  CCTTTGGCCG
40   1001  CCGTAACCTT  GGTCTTGCTT  CGCGAGGGCG  CGCAGAAATA  TTTTGCCGGC
     1051  AGTTTTTACC  GGGGCAGGTA  G

```

This corresponds to the amino acid sequence (SEQ ID NO: 790; ORF121ng-1):

```

      1  MYRRKGRGIK  PWMGAGAAFA  ALVWL VYALG  DTLTPFAVAA  VLAYVLDPLV
     51  EWLQKGLNR  ASASMSVMVF  SLILL LALLL  IIVPMLVGQF  NNLASRLPQL
45    101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  FQAHTGELSN  ALKAWFPVLM
     151  KQGGNIVSSI  GNLLLPPLLL  YYFLLDWQRW  SCGIAKLVR  RFAGAYTRIT
     201  GNLNELGEF  LRGQLLVMLI  MGLVYGLGLM  LVGLDSGFAI  GMVAGILVFV
     251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVW  AVFAVGQFLE  SFFITPKIVG
     301  DRIGLSPFWV  IFSLMAFGEL  MGFVGMLAGL  PLAAVTLVLL  REGAQKYFAG
50    351  SFYRGR*

```



ORF121ng-1 (SEQ ID NO: 790) and ORF121-1 (SEQ ID NO: 784) show 97.5% identity in 356 aa overlap:

```

5      10      20      30      40      50      60
orf121-1.pep MYRRKGRGIKPWMGAGAAFAALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR
orf121ng-1 MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60

10     70     80     90    100    110    120
orf121-1.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
orf121ng-1 ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      70     80     90    100    110    120

15    130    140    150    160    170    180
orf121-1.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPPLLYYFLLDWQRW
orf121ng-1 EIDQASIIAWFQAHTGELSNAKAWFPVLMKQGGNIVSSIGNLLLLPPLLYYFLLDWQRW
      130    140    150    160    170    180

20    190    200    210    220    230    240
orf121-1.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
orf121ng-1 SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI
      190    200    210    220    230    240

25    250    260    270    280    290    300
orf121-1.pep GMLAGILVFVPYLGAF TGLLLATVAALLQFGSWGILSVWAVFAVGQFLESFFITPKIVG
orf121ng-1 GMVAGILVFVPYLGAF TGLLLATVAALLQFGSWGILAVWAVFAVGQFLESFFITPKIVG
      250    260    270    280    290    300

30    310    320    330    340    350
orf121-1.pep DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
orf121ng-1 DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX
      310    320    330    340    350
```

In addition, ORF121ng-1 (SEQ ID NO: 790) shows homology to a permease (SEQ ID NO: 1162) from *H.influenzae*:

```

35      sp|P43969|PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
      Score = 69.9 bits (168), Expect = 2e-11
      Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)

40      Query: 26  VYALGDTLTPFAVAAYVLDPLVEWL-QKKGLNRRASASMSVMVFSXXXXXXXXXXXXV 84
      +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
      Sbjct: 32  IYFFGDLIAPLLIALVLSYLLEIPINFLNQYLKCPRLATILIFGSFIGLAAVFFLVLP 91

      Query: 85  MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNAK 143
      ML Q +L S LP + N WL N Y E ID + + + F + ++ +
      Sbjct: 92  MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFGE 147
45      Query: 144 AWFVPVLMKQGGNIVSSIGNXXXXXXXXXXXXDWQRWSCGIAKLVPRRFAGAYTRITGNL 203
```

```

      +   +   +   N+VS                      D      G+++ +P+   A+ R   +
Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRLFKNRNLAFXRWK-EM 206

Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXSGFAIGMVAGILVFVPYXXXXXXXXXXXX 263
      +   +   ++ G+                      +   +   G+ V VPY
5 Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPLLLAFVGLSVLVPYIGAVIVTIPVA 266

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
      QFG           +   FAV Q L+   + P +   + L P   +I S++ FG L GF
Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

Query: 324 VGMLAGLPLAAVTLVLL 340
      G+   +PLA +   ++
10 Sbjct: 327 WGVFFAIPLATLVKAVI 343

```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 94

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 791):

```

20 1  ..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
    51  TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
    101 TTTGCACGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
    151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
    201 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
    251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTCTGTGTC
25 301 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
    351 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACCTCT
    401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
    451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTG
30 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAG..

```

This corresponds to the amino acid sequence (SEQ ID NO: 792; ORF122):

```

35 1  ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR
    51  LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGFLC
    101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCRT
    151 EQRVGNVQQQ RIGIGVSEQP FFKWDFNSAK YQ..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 793):

```

40 1  ATATCGTACT GGGCAAGCAG TTCGCCGGAT TTTTGGGAAG TAGATACCGC
    51  GCCTTTGATT TTTTGGCCG TCTTACCAA GGCTTCGATG AAAAAGTTGA
    101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTGC
    151 ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGTG TCGTCATATT
    201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
    251 TTTGCACGTC CTGCCGCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC

```

5  
 10  
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG  
 351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 401 ATGTTGGCAC GCATTGCGG AATGTGCGGC GCGAGTTTGG GTTTCTGTGC  
 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC  
 501 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT  
 551 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 601 GAGCAGCGCG TCGGTAACGG CGTGACGAG CGCATCGGCA TCGGAGTGTC  
 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAT CTCCGCCAAG TATCAGCTTT  
 701 CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT  
 751 CGTCATCGTT TGTGTTCTCTG A

This corresponds to the amino acid sequence (SEQ ID NO: 794; ORF122-1):

15  
 1 1SYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS  
 51 TAFSAAMRLS SSCVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR  
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYQA IGGDVGTHLR NVREFGFLC  
 151 NHGRIDIDL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT  
 201 EQRVGNVQQR RIGIGVSEQP PFKWDFNSAK YQLSAFGQLV DIVALSDTDV  
 251 RHRLCS\*

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF122 (SEQ ID NO: 792) shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) (SEQ ID NO: 796) from strain A of *N. meningitidis*:

25  
 orf122.pep TAFSAALRLSPSXLVIFLSFGKPYQQTAAI  
 orf122a FLPLLPKASMKKLMVEVPVPMPIYSFSGTNSTAFSAAMRLSSSCVIFLSFGKPYQQTAAI  
 30  
 orf122.pep LTFFCTSCPPRSNAYQQYRRRLRLYAFHPPEIAEFFVGFAFDVDARNVYQAIGGDVGTHLR  
 orf122a LTFFXTSCPPRSNPYQQYRRRLRLYAFHAPETIEFFVGFAFXVDARNVYQAIGGDVGTHLR  
 35  
 orf122.pep NVRRECGFLCNHGRIDIDLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT  
 orf122a NMRREFGFLCNHGRIDIDLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT  
 40  
 orf122.pep EQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQ  
 orf122a EQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLVDIVALSDTDVRHRLCSX

45 The complete length ORF122a nucleotide sequence (SEQ ID NO: 795) is:

1 ATATCATATT GGGCAAGCAG TTCACTGGAT TTTTGGGAAG TAGATACCGC

51 GCCTTTGATT TTTTGGCCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA  
101 TGGTCGAACC GGTACCGATG CCGATGTATT CGTTTTCGGG TACGAATTCG  
151 ACTGCNTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT  
201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
5 251 TTNNNACGTC CTGCCCGCCG CGTTCAAATC CTTACCAGCA ATACCGCCGC  
301 CTGCGACTCT ATGCCTTCCA TGCGCCCGAG ATAACCGAGT TTTTCGTTGG  
351 TTTTGCCTTT GANGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
401 ATGTTGGCAC GCATTGCGG AATATGCGGC GCGAGTTTGG GTTCTGTGC  
10 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCTGCG GCCTGAACGC  
501 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACTCT  
551 GCGGCGGTGT CGGGGAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAGCTTT  
701 CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT  
15 751 CGTCATCGTT TGTGTTCTTG A

This encodes a protein having amino acid sequence (SEQ ID NO: 796):

1 ISYWASSSLD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PMYSFSGTNS  
20 51 TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFFXTSCPP RSNPYQQYRR  
101 LRLYAFHAFE ITEFFVGFAF XVDARNVYAQ IGGDVGTHLR NMRREFGFLC  
151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGVGEM AADIAQTCRT  
201 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDTDV  
251 RHRLCS\*

25 ORF122a (SEQ ID NO: 796) and ORF122-1 (SEQ ID NO: 794) show 96.9% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122a.pep	ISYWASSSLDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPMYSFSGTNSTAFSAAMRLS					
30	orf122-1	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
	orf122a.pep	SSCVVIFLSFGKPYQQTAAILTFFXTSCPPRSNPYQQYRRRLRYAFHAFEITEFFVGFAF					
35	orf122-1	SSCVVIFLSFGKPYQQTAAILTFFCTSCPPRSNAYQQYRRRLRYAFHPPEIAEFFVGFAF					
		70	80	90	100	110	120
	orf122a.pep	XVDARNVYAQIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
40	orf122-1	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
	orf122a.pep	FELCGGVGEMAAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLV					
45	orf122-1	FELCGGVGEMAAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
	orf122a.pep	DIVALSDTDVHRRLCSX					
50	orf122-1	DIVALSDTDVHRRLCSX					
		250					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 (SEQ ID NO: 792) shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) (SEQ ID NO: 798) from *N.gonorrhoeae*:

```

5      orf122.pep                                TAFSAALRLSPSXLVIFLSFGKPYQQTAAI    30
      orf122ng      FLPLLPKASMKKLMVEPVPMYFSFGTNSTAFSAAMRLSSSCVVI FLFSFGKPYQQTAAI    80
      orf122.pep      LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR    90
      orf122ng      LTFFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR    140
10     orf122.pep      NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCTRT    150
      orf122ng      NVRCEFGFLCNHGRIDIDHLPTLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCTRT    200
      orf122.pep      EQRVGNVGVQQRIGIGVSEQPFFKWDFNSAKYQ                                182
15     orf122ng      EQRVGNVGVQQRVGIRMPEQPFFKWDFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS    256

```

The complete length ORF122ng nucleotide sequence (SEQ ID NO: 797) is:

```

20     1  ATGTCGTACC GGGCAAGCAG TTCGCCGAT TTTTGGAGG TTGAAACCGC
      51  GCCTTTGATT TTTTACCGC TTTTGCCAA GGCTTCGATG AAGAAATTGa
      101 tgGTCGAACC GgtaCCGATG CCGATGTATT CGTTTTCGGG TACGAATTG
      151 ACTGCTTTT CGGCGGCGAT GCGCttgAgt TCgtcttgcg TcgTCATATT
      201 TTTAtccttt gGGAAccct atcaAcaAAc agccgccatC TTAACATTTT
      251 TTTGCACGtc ctggccgccg cgttcaAATc cgtaccaGca ataccgccgc
      301 ctgcgctctCT AtgcCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
25     351 TTTTGCTTT GATatTGACG CACGAAATAT CGatacCCAa atcggcgGCG
      401 ATGTTGGCAC GCATTTGCGG AATGTGCGGT GCGAGTTTGG GTTTCTGTGC
      451 AATCACGTC GTATCGACAT TGACCACCTG CCAACCCTGC GCCTGAACGC
      501 TTTGATACGC CGCACGCAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
      551 GCGGCGGTGT CGGGAATATG GCTGCCGATG TCGCCAAAC CTGCCGACC
30     601 GAGCAGCgcg tcggtaaCGG CGTGCAGCAG cgcgTcgGCA TCCGAATGCC
      651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCCGCCAAG TATCAGCTTT
      701 CTGCCTTCGG TCAATTGGTG GACATCGTAG CCCTGTCCGA TACGGATATT
      751 CGTCATCGTT TGTGTTCTG A

```

35 This encodes a protein having amino acid sequence (SEQ ID NO: 798):

```

40     1  MSYRASSPD FLEVETAPLI FLPLLPKASM KKLMEPVPM PMYFSGTNS
      51  TAFSAAMRLS SSCVVI FLFSFGKPYQQTAAI LTFFCTSWPP RSNPYQQYRR
      101 LRLYAFHPPE IAEFFVGFAF DIDARNIDTQ IGGDVGTHLR NVRCFEGFLC
      151 NHGRIDIDL PTLRLNALIR RTQKDAAVRI FELCGVGKM AADVAQTCTRT
      201 EQRVGNVGVQ RVGIRMPEQP FFKWDFNSAK YQLSAFGQLV DIVALSDTDI
      251 RHRLCS*

```

ORF122ng (SEQ ID NO: 798) and ORF122-1 (SEQ ID NO: 794) show 92.6% identity in 256 aa overlap:

-560-

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
	orf122-1.pep	70	80	90	100	110	120
	orf122ng	SSCVVIFLSFGKPYQQTAAILTFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
10		70	80	90	100	110	120
	orf122-1.pep	130	140	150	160	170	180
	orf122ng	DVDARNVYAIQGGDVGTHLRNVREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
15		130	140	150	160	170	180
	orf122-1.pep	190	200	210	220	230	240
	orf122ng	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFNSAKYQLSAFGQLV					
20		190	200	210	220	230	240
	orf122-1.pep	250	DIVALSDTDVRHRLCSX				
	orf122ng	250	DIVALSDTDIRHRLCSX				
25		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 95

30 The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 799):

35

```

1  ..GCCGGCGCGA GTGCGAACAA CATTTCGCG CGTTTTCGCG AAACACCCGT
51  CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCC
101 TTACCGAATA TGAAACTTC CTGCTGCTTA TCGGCTCGGT ATTTGCGCCG
151 ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAAC GGCGTGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 800; ORF125):

```

1  ..AGASANNISA RFAETPVAVS VTLLIGTVLAV MLPVTEYENF LLLIGSVFAP
51  MGGFDCRLFR LETA*

```

40 Further work revealed the complete nucleotide sequence (SEQ ID NO: 801):

```

1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCTCCGCCA TCGGGCTGAT
51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
101 TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CGGCTCTACT TTTGGGTGAT
151 GCCGTCGGCG GCGCGCTGTT TTTTGCGGCG GCGTATATCG GCGCACTGAC

```

5  
 10  
 15  
 20

```

201 CGGACGCAGC TCGATGGAAA GCGTGCGCCT GTCGTTCCGGC AAACGCGGTT
251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
301 GTGATGATTT ACGCCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGT
351 GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GGCGCGCTGA
401 TTGTGCTGTG GCTGGTTTTT GCGCACGCA AAACAGGCGG GCTGAAAACC
451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAAGT
501 CTTTTCACG GCAGGCAGCA CCGCCGCACA GGTTCAGAC GGCATGAGTT
551 TCGGAACGGC AGTCGAGCTG TCCGCCGTGA TGCCGCTTTC CTGGCTGCCG
601 CTTGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCTT
651 GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG
701 GTTTGGCAGC GCGGTTGTTC ACCGAGAGAA CCGACGTGGC AAAAATCCTG
751 CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTGC TCCTCTCCAC
801 CGTTACCACA ACCTTTCTCG ATGCCTATTC CGCCGGCGCG AGTGCGAACA
851 ACATTTCCGC GCGTTTTGCG GAAACACCCG TCGCTGTCGG CGTTACCCCTG
901 ATCGGCACGG TACTTGCCGT CATGCTGCCC GTTACCGAAT ATGAAAACCTT
951 CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGGCG GTTTTGATTG
1001 CCGACTTTTT CGTCTTGAAA CGGCGTGAGG AGATTGAAGG CTTTGACTTT
1051 GCCGGACTGG TTCTGTGGCT TCGGGGCTTC ATCCTCTACC GCTTCCTGCT
1101 CTCGTCCGGC TGGGAAAGCA GCATCGGTCT GACCGCCCCC GTAATGTCTG
1151 CCGTTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTTAA AAAAACCCTA
1201 TCTTTACAAA GGAACCCGTC ATGA
  
```

This corresponds to the amino acid sequence (SEQ ID NO: 802; ORF125-1):

25  
 30

```

1 MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSVA NMLQLAGWTA
101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
151 VSMLLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP
201 LAADYTRHAR RPFATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL
251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL
301 IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF
351 AGLVLWLAGF ILYRFLSSG WESSIGLTAP VMSAVAIATV SVRLFFKKQT
401 SLQRNPS*
  
```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 (SEQ ID NO: 800) shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) (SEQ ID NO: 804) from strain A of *N. meningitidis*:

40  
 45

```

orfl25.pep                                10      20      30
                                AGASANNISARFAETPVAVSVTLIGTVLAV
                                ||:|||||:::| |:|:|:::|:|
orfl25a      KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVVGTLLAV
              250      260      270      280      290      300

              40      50      60
orfl25.pep    MLPVTEYENFLLIGSVFAPMGGFDCRLFRLETAX
              :|||||:
orfl25a      LLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG
              310      320      330      340
  
```

The ORF125a partial nucleotide sequence (SEQ ID NO: 803) is:

1 ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCCGCCA TCGGGCTGAT  
 51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACACTGC  
 101 TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CNGCTCTGCT TTTGGGTCAT  
 151 GCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC  
 5 201 CGGACNCANC TCGATGGAAA GCGTGCGCCT GTCGTTCGGC AAACGCGGTT  
 251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG  
 301 GTGATGATTT ACGCCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT  
 351 GTGGGACGGC GAATCTTTTG TCTGGTGGG ATTGGCAAAC GCGCGCTGA  
 10 401 TTGTGCTGTG GCTGGTTTTT GCGCACGCA AAACAGGCGG GCTGAAAACC  
 451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAANT  
 501 NTTTTCCACG GCAGGCAGCA CCGCCGANN GGTNNCAGAC GGCATGAGTT  
 551 TCGGAACGGC AGTCGAGCTG TCCGCCGTNA TGCCGCTTTC TTGGCTGCGG  
 601 CTGGCCGCGG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCCCT  
 651 GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG  
 15 701 GTTTGGCAGC GCGGTTGTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG  
 751 CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTG TCCTGTCGAC  
 801 CGTTACCACC ACTTTTCTCG ATGCNTACTC CGCCGGCGTA AGTGCCAACA  
 851 ATATTTCCGC CAAACTTTTC GAAATACCNA TCGCCGTTGC CGTCGCCGTT  
 901 GTCGGCACAC TGCTTGCCGT CCTCCTGCCC GTTACCGAAT ATGAAAACCT  
 20 951 CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGGCG GTTTTGATTG  
 1001 CCGACTTTTT CGTCTTGAAG CCGCGTGAGG AGATTGAAGG C..

This encodes a protein having the partial amino acid sequence (SEQ ID NO: 804):

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLWQ RGLAALLLGH  
 25 51 AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSVA NMLQLAGWTA  
 101 VMYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPFATLTAT LAYTLTGWM YALGLAAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPIAVAVAV  
 30 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a (SEQ ID NO: 804) and ORF125-1 (SEQ ID NO: 802) show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
35	orf125a.pep	MSGNASSXSSSAAIGLIWFGAAVSVIAEISTGTLLAPLWQ RGLAALLLGHAVGGALFFAA					
	orf125-1	MSGNASSPSSSSAIGLIWFGAAVSVIAEISTGTLLAPLWQ RGLAALLLGHAVGGALFFAA					
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf125a.pep	AYIGALTGXSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
	orf125-1	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
		70	80	90	100	110	120
		130	140	150	160	170	180
45	orf125a.pep	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEXFSTAGSTAAXVD					
	orf125-1	ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAQVSD					
		130	140	150	160	170	180
		190	200	210	220	230	240
50	orf125a.pep	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAAALF					
	orf125-1	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMYALGLAAALF					



-563-

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf125a.pep	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAV					
	orf125-1	TGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVTL					
		250	260	270	280	290	300
		310	320	330	340		
	orf125a.pep	VGTLLAVLLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG					
10	orf125-1	IGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVWLAGF					
		310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

ORF125 (SEQ ID NO: 800) shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) (SEQ ID NO: 806) from *N.gonorrhoeae*:

15	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
	orf125.pep	MLPVTEYENFLLIGSVFAPM-GGFDCRLFRLETA	64
20	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGFDCRLFCLKTA	343

An ORF125ng nucleotide sequence (SEQ ID NO: 805) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 806):

25	1	MSGNASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLLAPLGWQ	RGLAALLLGH
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA
	101	VMIYVGATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT
	151	VSMLLMLLAV	LWLSVEVFAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL
	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCW	MYALGLAAAL	FTGETDVAKI
	251	LLGAGLGITG	ILAVVLSTVT	TTFLDTYSAG	ASANNISARF	AEIPVAVGV
30	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGFDCRLFCL	KTA*

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 807):

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCGCCGCCA	TCGGGCTGGT
35	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGTGC
	101	TCGCCCCCTT	GGGCTGGCAG	CGCGGTCTGG	CGGCCCTGCT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTCGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGAAA	GTGTGCGCCT	GTCGTTCCGC	AAATGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTT	ACGTCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
40	351	GTGGGACGGC	GAATCCTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCACTGA
	401	TCGTGCTGTG	GCTGGTTTTT	GGCGCACGCA	GAACGGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GCTTGCCGTG	TTGTGGTTGA	GCGTCGAAGT
	501	GTTTCGCTTC	TCCGGCACAA	ACGCCGCGCC	CGCCGTTTCA	GACGGCATGA
	551	CCTTCGGAAC	GGCAGTCGAA	CTGTCCGCCG	TCATGCCGCT	TTCTGGCTG
45	601	CCGCTGGCCG	GCGACTACAC	GCGCCAAGCA	CGCGCCCGT	TTGCGGCAAC
	651	CCTGACGGCA	ACGCTCGCCT	ATACGCTGAC	GGGCTGCTGG	ATGTATGCCT

-564-

5  
10

```

701 TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GGCGAAAATC
751 CTGTTGGGCG CGGGCTTGGG CATAACGGGC ATTCTGGCAG TCGTCCTCTC
801 CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGC GA
851 ACAACATTTT CCGCGGTTTT GCGGAAATAC CCGTCGCTGT CCGCGTTACC
901 CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
951 CTTCTGCTG CTATCGGCT CGGTATTTGC GCCGATGGCG GCGGTTTTGA
1001 TTGCCGACTT TTTCGTCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
1051 TTTGCCGGAC TGGTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
1101 GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
1201 CAATCTTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence (SEQ ID NO: 808; ORF125ng-1):

15  
20

```

1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
151 VSMMLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
201 PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAAL FTGETDVAKI
251 LLGAGLGITG ILAVVLSTVT TFLDTSAG ASANNISARF AEIPVAVGV
301 LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
351 FAGLVWLAG FILYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFPPKKT
401 QSLQRNPS*

```

25 overlap:

30  
35  
40  
45  
50

```

              10      20      30      40      50      60
orf125-1.pep  MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLAPLGWQRGLAALLLGHAVGGALFFAA
              |||||:||||:|||||
orf125ng-1    MSGNASSPSSSSAIGLVWFGAAVSIAEISTGTLAPLGWQRGLAALLLGHAVGGALFFAA
              10      20      30      40      50      60

              70      80      90      100     110     120
orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
              |||||:|||||
orf125ng-1    AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG
              70      80      90      100     110     120

              130     140     150     160     170     179
orf125-1.pep  ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFASTAGSTAAQ-VS
              |||||:|||||
orf125ng-1    ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSTNAAPAVS
              130     140     150     160     170     180

              180     190     200     210     220     230     239
orf125-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCWMYALGLAAAL
              |||:|||||
orf125ng-1    DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCWMYALGLAAAL
              190     200     210     220     230     240

              240     250     260     270     280     290     299
orf125-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGV
              |||||:|||||
orf125ng-1    FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDTSAGASANNISARFAEIPVAVGV
              250     260     270     280     290     300

```

-565-

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVTEYENFLL	LIGSVFAPMAAVLIADFFVLKRREEIEG	FDFAGLVWL	LAG			
5	orf125ng-1	LIGTVLAVMLPVTEYKNFLL	LIGSVFAPMAAVLIADFFVLKRREEIEG	FDFAGLVWL	LAG			
		310	320	330	340	350	360	
		360	370	380	390	400		
	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
10	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 96

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 809):

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
20	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.ACGGA
25	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

This corresponds to the amino acid sequence (SEQ ID NO: 810; ORF126):

	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCL	NHTMMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

Further work revealed the complete nucleotide sequence (SEQ ID NO: 811):

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TTGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
40	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
45	451	CTCGACGGGC	GGCAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCCTGCAAG

5	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGAGCA	CACCAGCACC	CTGCGGGCA	TACGCGGCGA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCCTC
	701	TGCTCCATCG	GCCTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACCACGTC
	751	TTCGTCATCG	GCGCGACCCA	AATCGAAAGC	GAAGCGCAAG	CCCCCGCCAG
10	801	CTGCGCTTCA	GGGTTGGAAC	TCTTGTCCGC	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
	901	CTCAACCACC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCTGAT
	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCCGTAA
	1001	CCGCCGCCGC	CGCCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence (SEQ ID NO: 812; ORF126-1):

15	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKGCRERGEHA	AAVVAAML
	51	PAEEAVEATP	EVVRLGRQSI	PLWRGIRCL	NTHMMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGGVADDEI	VRWRADDIAE	REPQLGGRFS	DGIYLPTEGQ
	151	LDGRQILSAL	ADALDELNVP	CHWEHECVPE	GLQAQYDWLI	DCRYGAKTA
	201	WNQSPEHTST	LRGIRGEVAR	VYTPEITLNR	PVRLHPRYP	LYIAPKENHV
20	251	FVIGATQIES	ESQAPASVRS	GLELLSALYA	IHPAFGEADI	LEIATGLRPT
	301	LNHHNPEIRY	NRARLIEIN	GLFRHGFMI	<u>PAVTA</u> AAARL	<u>AVALF</u> DGKDA
	351	PERDKESGLA	YIRROD*			

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 (SEQ ID NO: 810) shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) (SEQ ID NO: 814) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
30	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP					
	orf126a	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
		70	80	90	100	110	120
35	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHTMQENGSLIVVHGGQDKPLSSEFVRHLKRGGXTDDEI					
	orf126a	EVVRLGRQXIPLWRGIRCHLKTAMPXENGSLIVVHGGQDKPLSNEFVRHLKRGGVADDXI					
		70	80	90	100	110	120
		130	140	150	160	170	180
40	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE					
	orf126a	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE					
		130	140	150	160	170	180

The complete length ORF126a nucleotide sequence (SEQ ID NO: 813) is:

45           1   ATGACCCGTA   TCGCCATCCT   CGGCGGCGGC   CTCTCNGGAA   GGCTGACCGC  
          51   ACTGCAGCTT   GCAGAAACAAG   GTTATCAGAT   TGCACCTTTTC   GATAAAGGCT  
         101   GCCGCCGGGG   CGAACACGCC   GCCGCCTATG   TTGCCGCCGC   CATGCTCGCG

5  
10  
15  
20

```

151 CCTGCGGCGG AAGCGGTCGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 GCAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACNAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAAACGCG
601 TGGAACCAAT CCCCCGANNNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
751 TTCGTATCTG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCGTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
901 CTCAATCACC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGCG TATATCCGAA GACAAGATTA
1101 A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 814):

25  
30

```

1 MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51 PAEEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
201 WNQSPXXTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV
251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA
351 PERDEESGLA YIRRQD*

```

ORF126a (SEQ ID NO: 814) and ORF126-1 (SEQ ID NO: 812) show 95.4% identity in 366 aa overlap:

35  
40  
45  
50

```

          10      20      30      40      50      60
orf126a.pep MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAYVAAAMLAPAAEAVEATP
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf126-1    MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAYVAAAMLAPAAEAVEATP
          10      20      30      40      50      60

          70      80      90      100     110     120
orf126a.pep EVVRLGRQXIPLWRGIRCHLKTTPAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf126-1    EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
          70      80      90      100     110     120

          130     140     150     160     170     180
orf126a.pep VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf126-1    VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE
          130     140     150     160     170     180

          190     200     210     220     230     240
orf126a.pep DLQAQYDWLIDCRGYGAKTAWNQSPXXTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf126-1    GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP

```

-568-

		190	200	210	220	230	240
		250	260	270	280	290	300
	orf126a.pep	LYIAPKENXV FVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIATGLRPT					
5	orf126-1	LYIAPKENHV FVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf126a.pep	LNHHNPEIRYNRARRLIEINGLFRHGFMISSPAVTAAAVRLAVALFDGKXAPERDEESGLA					
10	orf126-1	LNHHNPEIRYNRARRLIEINGLFRHGFMISSPAVTAARLAVALFDGKDAPERDKESGLA					
		310	320	330	340	350	360
	orf126a.pep	YIRRODX					
15	orf126-1	YIRRODX					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 (SEQ ID NO: 810) shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) (SEQ ID NO: 816) from *N.gonorrhoeae*:

20	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHAAAYVAAAMLAPAAEAVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
	orf126ng	EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
25	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ	180

An ORF126ng nucleotide sequence (SEQ ID NO: 815) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 816):

	1	MTRIAVLGGG	LSGRLTALQL	AEQGYQIELF	DKGTRQGEHA	AAYVAAAMLA
	51	PAAEAVEATP	EVIRLGRQSI	PLWRGIRCL	NTLTMMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGGVADDEI	VRWRADEIAE	REPQLGGRFS	DGIYLPTEGQ
35	151	LDGRQILSAL	ADALDELNVP	CHWEHECAPQ	DLQAQYDWVI	DCRGYGAKTA
	201	WNQSPEHTST	LRGIRGEVRG	FTRPKSRSTA	PCACCTRAIR	STSPRKKTTS
	251	SSSARPKSKA	KAKPPPAYVP	GWNSYPRSMP	STPPSAKPTS	SKWRPGLRPT
	301	LNHHNPEIRY	SRERRLIEIN	GLFRHGFMISS	PAVTAAAVRL	AVALFDGKDA
	351	PERDEESGLA	YIGRQD*			

40 Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 817):

	1	ATGACCCGTA	TCGCCGTCCT	CGGAGGCGGC	CTTTCCGGAA	GGCTGACCGC
	51	ATTGCAGCTT	GCAACAACAG	GTTATCAGAT	TGAACCTTTC	GACAAGGGCA
	101	CCCGCCAAGG	CGAACACGCC	GCCGCCTATG	TTGCCGCCGC	GATGCTCGCG

5  
10  
15  
20

```
151 CCTGCGGCGG AAGCGGTCGA GGCAACGCCC GAAGTCATCA GGCTGGGCAG
201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA
251 CGATGATGCA GGAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACGAAATC GTCCGTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
501 GAACGTCCCT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG
551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG
601 TGGAAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA
651 AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC
751 TTCGTATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCAGCCAG
801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG
851 CTTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG
901 CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT
951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTCC CCCGCCGTAA
1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
1051 CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
1101 A
```

This corresponds to the amino acid sequence (SEQ ID NO: 818; ORF126ng-1):

25  
30

```
1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA
51 PAEEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA
201 WNQSPEHTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
301 LNHHNPEIRY SRERLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
351 PERDEESGLA YIGRQD*
```

ORF126ng-1 (SEQ ID NO: 818) and ORF126-1 (SEQ ID NO: 812) show 95.1% identity in 366 aa overlap:

35  
40  
45  
50

```
10 20 30 40 50 60
orf126-1.pep MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP
|||:|||||||||||||||||||||||:|||||||||||||||||
orf126ng-1 MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHAAAYVAAAMLAPAAEAVEATP
10 20 30 40 50 60

70 80 90 100 110 120
orf126-1.pep EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
||:|||||||||||||||||||||||||||||||||||||||||
orf126ng-1 EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
70 80 90 100 110 120

130 140 150 160 170 180
orf126-1.pep VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE
|||:|||||||||||||||||||||||||||||||||||||||:|
orf126ng-1 VRWRADEIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ
130 140 150 160 170 180

190 200 210 220 230 240
orf126-1.pep GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPEITLNRPVRLHPRYP
|||:|||||||||||||||||||||||||||||||||||||||
orf126ng-1 DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPEITLNRPVRLHPRYP
```

-570-

		190	200	210	220	230	240
		250	260	270	280	290	300
5	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT					
	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPT					
		250	260	270	280	290	300
		310	320	330	340	350	360
10	orf126-1.pep	LNHHNPEIRYNRARRLIEINGLFRHGFMISSPAVTAARLAVALFDGKDAPERDKESGLA					
	orf126ng-1	LNHHNPEIRYSRERRLIEINGLFRHGFMISSPAVTAARLAVALFDGKDAPERDEESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 (SEQ ID NO: 818) shows homology to a putative *Rhizobium* oxidase flavoprotein (SEQ ID NO: 1163):

20	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]	
	Length = 327	
	Score = 169 bits (423), Expect = 3e-41	
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)	
25	Query: 3	RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQGEHXXXXXXXXXXXXXXXXXXXXXXXXX 62
		RI V G G++G A QL G+++ L ++ G
	Sbjct: 2	RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEPEV 60
	Query: 63	IRLGRQSIPLWRGIRCLNLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
		+ LGR + W + G+L+V G+D F R G DE+
	Sbjct: 61	LTLGRLAADWEEA-----LPGHVHRRGTLVAVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
30	Query: 123	WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
		IA EP L GRF ++ E LD RQ L+ALA L++ + +
	Sbjct: 114	-----IAALEPDLAGRFRRALFFRQEAHLDPQALALAAGLEDARMRLTLG---VVGES 165
	Query: 183	QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYPLY 242
		+D V+DC G LRG+RGE+ V T E++L+RPVRLHPR+P+Y
	Sbjct: 166	DVDHDRVVDCTGAA-----QIGRLPGLRGVGMELCVETTEVSLSRPVRLHPRHPIY 218
35	Query: 243	IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
		I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP
	Sbjct: 219	IVPRDKNRFMVGATMIESDDGGPITARSLMELLNAAAYAMHPAFGEARVTETGAGVRPAYP 278
	Query: 303	HHNPEIRYSRERRLIEINGLFRHGFMISS 331
		+ P R ++E R + +NGL+RHGF+++P
40	Sbjct: 279	DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 97



The following DNA sequence, believed to be complete, was identified in *N.meningitidis* (SEQ ID NO: 819):

```

      1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
      51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
5      101  TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
      151  CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GGGAGGTTTA AACAAACATC
      201  TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
      251  GTTTGAATGG AATCGTCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
      301  AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA
10     351  TGAAGATCTA GTAACCTTTA aTTTGCAAGA AGTCCGCCAG TTCGTGTAGT
      401  GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
      451  GTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 820; ORF127):

```

15      1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
      51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
      101  KAVAIDKDKN PFIKMNENL VTFICKKSAS SCSDGLDYFK GNDKDKLLK
      151  *

```

20 Further work revealed the following DNA sequence (SEQ ID NO: 821):

```

      1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
      51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
25     101  TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
      151  CATTTTATGG AAAAGTTTTA TCTGCAGAAAT GGGAGGTTTA AACAAACATC
      201  TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
      251  GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
      301  GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
      351  AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
30     401  GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 822; ORF127-1):

```

35      1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
      51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
      101  AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 (SEQ ID NO: 820) shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) (SEQ ID NO: 824) from strain A of *N. meningitidis*:

```

40      orf127.pep      10      20      30      40      50      60
      MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFYLQN
      |||||
      orf127a         10      20      30      40      50      60
      MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLEN AHFMEKFYLQN

```

-572-

5  
10

	70	80	90	100	110	120
orf127.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIVARXALDSKFMLKAVAIDKDKNPFIKMNENL					
orf127a	GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFIKMNENL					
	70	80	90	100	110	
	130	140	150			
orf127.pep	VTFICKKSASSCSDGLDYFKGNDKDKLLKX					
orf127a	VTFICKKSASSCSDGLDYFKGNDKDKLLKX					
	120	130	140	150		

The complete length ORF127a nucleotide sequence (SEQ ID NO: 823) is:

15  
20

1	ATGACTGATA	ATCGGGGGTT	TACGCTGGTT	GAATTAATAT	CAGTGGTCTT
51	GATATTGTCT	GTACTTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG
101	TTGAGAAAGC	AAAGATAAAT	ACAGTGCGGG	CAGCCTTGTT	AGAAAATGCA
151	CATTTTATGG	AAAAGTTTTA	TCTGCAGAAT	GGGAGATTTA	AACAAACATC
201	TACCAAATGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTGTATCC
251	GTTTGAATGG	AATCGCGCGC	GGGGCCTTAG	ACAGTAAATT	CATGTTGAAG
301	GCGGTAGCCA	TAGATAAAGA	TAAAAATCCT	TTTATTATTA	AGATGAATGA
351	AAATCTAGTA	ACCTTTATTT	GCAAGAAGTC	CGCCAGTTCG	TGTAGTGACG
401	GGCTGGATTA	TTTTAAAGGA	AATGATAAGG	ACTGCAAGTT	ACTTAAGTAG

This encodes a protein having amino acid sequence (SEQ ID NO: 824):

25

1	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	TVRAALLEN
51	HFMEKFYLQN	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK
101	AVAIDKDKNP	FIIKMNENLV	TFICKKSASS	CSDGLDYFKG	NDKDKLLK*

ORF127a (SEQ ID NO: 824) and ORF127-1 (SEQ ID NO: 822) show 99.3% identity in 149 aa overlap:

30  
35  
40

	10	20	30	40	50	60
orf127a.pep	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	TVRAALLEN	AHFMEKFYLQN
orf127-1	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	AHFMEKFYLQN
	10	20	30	40	50	60
	70	80	90	100	110	120
orf127a.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGI	ARGALDSKFMLK	AVAIDKDKNPFIKMNENLV			
orf127-1	GRFKQTSTKWPSLPIKEAEGFCIRLNGI	ARGALDSKFMLK	AVAIDKDKNPFIKMNENLV			
	70	80	90	100	110	120
	130	140	150			
orf127a.pep	TFICKKSASSCSDGLDYFKGNDKDKLLKX					
orf127-1	TFICKKSASSCSDGLDYFKGNDKDKLLKX					
	130	140	150			

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 (SEQ ID NO: 820) shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) (SEQ ID NO: 826) from *N.gonorrhoeae*:

```

5      orf127.pep  MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN  60
      orf127ng    MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLENAHFMEKFYLQN  60

      orf127.pep  GRFKQTSTKWPSLPIKEAEGFCIRLNGIVARXALDSKFMLKAVAIDKDKNPFIKMNENL  120
      orf127ng    GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFIKMNENL  119

10     orf127.pep  VTFICKKSASSCSDGLDYFKGNDKDCKLLK  150
      orf127ng    VTFICKKSASSCSDRLDYFKGNDKDCKLLK  149

```

The complete length ORF127ng nucleotide sequence (SEQ ID NO: 825) is:

```

15      1  ATGACTGATA ATCGGGGGTT TACTACTGGTT GAATTAATAT CAGTGGTCTT
      51  GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
      101  TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
      151  CATT TTTATGG AAAAGTTT TCTGCAGAAT GGGAGATTTA AACAAACATC
      201  TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
      251  GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
20     301  GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
      351  AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
      401  GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence (SEQ ID NO: 826):

```

25      1  MTDNRGFTLV ELISVVLILS VLALIVPSY RNYVEKAKIN AVRAAFLENA
      51  HFMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
      101  AVAIDKDKNP FIIKMENLV TFICKKSASS CSDRLDYFKG NDKDCKLLK*

```

ORF127ng (SEQ ID NO: 826) and ORF127-1 (SEQ ID NO: 822) show 100.0% identity in 149 aa overlap:

```

35      orf127-1.pep  MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
      orf127ng-1      MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN

      orf127-1.pep  GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV
      orf127ng-1      GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV

40      orf127-1.pep  TFICKKSASSCSDGLDYFKGNDKDCKLLKX
      orf127ng-1      TFICKKSASSCSDGLDYFKGNDKDCKLLKX

45      orf127-1.pep  TFICKKSASSCSDGLDYFKGNDKDCKLLKX
      orf127ng-1      TFICKKSASSCSDGLDYFKGNDKDCKLLKX

```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 98

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 827)

```

1  ..GTGTCGCTGG CTTCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATT
51  CAACCAAATG CGGAAAACCC GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT
101 ATTTATCTGG GGTTCAGCA GGGGTATTTC GATTGTAGTG CCGACGAGAA
151 CCCCGTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGTATCCCCT TTTGCTGATA TTTGCTGCA AAAAAACCAA ATCGTACGG
251 GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
301 GTTTTGGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCCAATACTT
351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTTGGCAGG TTCGCTGCTG
15 401 GCGGTTTACG GGCAAACGCA AAACGGCAGA CGGCAAACAG CAAATGGAAA
451 ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
551 CCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
601 TCCGACCCGC ATCCTGTCGG CAAGCCCCAT CGTATTTGTC GGCAAATCT
20 651 CTTATTCCCT ATACCTGTAC CATTGGATTT TTATTGCTTT CGCTCCGCTC
701 ATTAGAGGCG GGAAACAGCT CGGACTGCCT GCCG..

```

This corresponds to the amino acid sequence (SEQ ID NO: 828; ORF128):

```

1  ..VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFFQ GYFDLSADEN
25 51  PVLHIWSLAV EEQYLLLYPL LLIFCCKKTK SLRVLRNISI ILFLILTASS
101 FLPSGFYTDI LNQPNTYYLS TLRFPPELLAG SLLAVYGQTQ NGRRQTANGK
151 RQLLSSLCFG ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VFGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

```

30 Further work revealed the complete nucleotide sequence (SEQ ID NO: 829):

```

1  ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC
51  CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
151 GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
35 201 TTATACCCGC AGGATTAAGC GGATTATCC TGCCTTTATT GCGGCCGTGT
251 CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTCAAC
301 CAAATGCGGA AAACCGTGGA GCTTCTGCG GTTTTCTTGT CCAATATTTA
351 TCTGGGGTTT CAGCAGGGGT ATTTGATTT GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
40 451 CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
501 GCGTAACATC AGCATCATCC TGTTTTGTAT TTTGACTGCC TCATCGTTTT
551 TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCAA TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
45 701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCTTG CCTGTTCTGT
751 ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
801 CCTGCTGACG GCACTGCTTA TCCGAGTAT GCAATACGGG ACACTTCCGA

```

5  
10  
15  
20

```

851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTG GATTTTATT GCTTTCGCCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
1001 CGGCCGGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC
1101 GTCCCTGATA CTTGTCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGTTG CCCGGCGCGC CCCTTGCTGC GGAAAATCAT
1201 TTTCCGAAA CCGTCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGAG
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAA GCCAAAATCC
1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
1351 AACCGTTAT GTCGAAAATA CCGGATGAA GTTGAAAAAG CCGAAGCCGT
1401 TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA
1451 GATTTGAAGC GCAATCCTTC CTAATACCCG GGTTCACAGC CCGATTGAGG
1501 GAAACGTC AAGGATAGC CGCCGTCAA CCCGTCTATG TTTTGTCAA
1551 CAACACATCA ATCAGCCGTT CGCCCTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCCGCAA CCAATATCTC CGCCCATTC AGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTGATT AAAGATATTC CCAATGTGCA
1701 TTGGGTGGAC GCACAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGGAAATCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA
1851 CGGCGGCGCA TTGCAGTAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 830; ORF128-1):

25  
30  
35

```

1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGAFLGVDI FFVISGFLIT
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTVELSA VFSLNIYLG FQQGYFDLSAD ENPVLHIWSL AVEEQYYLLY
151 PLLLIFCKK TKSLRVL RNI SIILFLILTA SSFLPSGFYT DILNQPNTYY
201 LSTLRFPELL AGSLLAVYQ TQNGRRQTAN GKRQLSSSLC FGALLACLFV
251 IDKHNPFIPG MTL L PCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLI SYYLIEQPLR
351 KRKMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHLRPL PGAPLAAENH
401 FPETVLTLD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVP RFEAQSF LIPGF PARFR
501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
551 KSNQAVFDLI KDIPNVHWD AQKYL PKNTV EIYGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKSSHGGA LQ*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723) (SEQ ID NO: 1164)

ORF128 (SEQ ID NO: 828) and HI0392 (SEQ ID NO: 1164) show 52% aa identity in 180aa overlap:

45

```

Orf128: 1 VSLASVIASQIFLYEDFNQMRKTVELSAVFLSN IYLG FQQGYFDLSADENPVLHIWSLAV 60
++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV
HI0392: 46 MALVSFIASAI F IYND FNKL RKTIELAIAFLSNFYLG LTQGYFDLSANENPVLHIWSLAV 105

Orf128: 61 EEQXXXXXXXXXIFCCKKTKSLRVL R NIS IILFLILTASSFLPSGFYTDILNQPNTYYLS 120
E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS
HI0392: 106 EGQYYLIFPLILILAYKKFREV KVLFIITLILFFILLATSFV SANFYKEVLHQPN IYYLS 165

```

Orf128: 121 TLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLCFGALLACLVIDKHNPFI PGMT 180  
 LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T  
 HI0392: 166 NLRFPELLVGSLLAIYHNLSN-KVQLSKOVNNILAILSTLLLFSCFLMNNNIAFI PGIT 224

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 5 ORF128 (SEQ ID NO: 828) shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) (SEQ ID NO: 832) from strain A of *N. meningitidis*:

	orf128.pep	VSLASVIASQIFLYEDFNQMKTVELSAVF	10	20	30			
10	orf128a	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMKTVELSAVF	60	70	80	90	100	110
	orf128.pep	LSNIYLGFQQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISI	40	50	60	70	80	90
15	orf128a	LSNIYLGFQQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISI	120	130	140	150	160	170
	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNITYYSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	100	110	120	130	140	150
20	orf128a	ILFLILTASSFLPSGFYTDILNQPNITYYSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	180	190	200	210	220	230
	orf128.pep	RQLSSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I	160	170	180	190	200	210
25	orf128a	RQLSSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I	240	250	260	270	280	290
	orf128.pep	VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA	220	230	240			
30	orf128a	VFVGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSSL SYYLIEQPLRK R	300	310	320	330	340	350
	orf128a	KMTFKKAFFCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLT LGDSH	360	370	380	390	400	410

- 35 The complete length ORF128a nucleotide sequence (SEQ ID NO: 831) is:

	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
	51	CGTGCTATCC	GTCATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
40	151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTTCTT	TCCGGGATTT
	201	TTATACCCGC	AGGATTAAAG	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAAATGCGGA	AAACCGTGGA	GCTTCTGCG	GTTTTCTTGT	CCAATATTTA
	351	TCTGGGGTTT	CAGCAGGGGT	ATTCGATT	GAGTGCACG	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
45	451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
	501	GCGTAACATC	AGCATCATCC	TATTCTGAT	TTTGACTGCC	ACATCGTTTT

-577-

5 551 TGCCAAGCGG GTTTTATACC GATATTCTCA ACCAACCCTAA TACTTATTAC  
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT  
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC  
701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCTTG CCTGTTCTGT  
751 ATTGACAAAC ACAATCCGTT TATCCCGGA ATGACCCTGC TCCTTCCCTG  
801 CCTGCTGACG GCACTGCTTA TCCGAGTAT GCAATACGGG ACACCTCCGA  
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT  
901 TCCCTATACC TGTACCATTG GATTTTATT GCTTTCGCCC ATTACATTAC  
10 951 AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA  
1001 CGGCCGGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA  
1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC  
1101 GTCCCTGATA CTTGTCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC  
1151 AGGAACACCT CCGCCCGTTG CCCGCGCGC CCCTTGCTGC GGAAAATCAT  
1201 TTTCCGAAA CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG  
15 1251 GGGGTTTCTG GATTATGTCG GCAGCCGGA AGGGTGAAA GCCAAAATCC  
1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC  
1351 AACCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT  
1401 TTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA  
1451 GATTTGAAGC GCAATCCTTC CTAATACCCG GGTTCACAGC CCGATTACAGG  
20 1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA CCCGTCTATG TTTTGTCAA  
1551 CAACACATCA ATCAGCCGTT CGCCCCTGAG GGAGGAAAAA TTGAAAAGAT  
1601 TTGCCGCAA CCAATATCTC CGCCCATTG AGGCTATGGG CGACATCGGC  
1651 AAGAGCAATC AGGCGGTCTT TGATTGATT AAAGATATT CCAATGTGCA  
25 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG  
1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT  
1801 TATATGGGGC GGGAATTCA CAAACACGAA CGCCTGCTTA AATCTTCTCG  
1851 CGACGGCGCA TTGCAGTAG

This encodes a protein having amino acid sequence (SEQ ID NO: 832):

30 1 MQAVRYRPEI DGLRAVAVLS VMIFHLNLRW LPGAFLGVDI FFVISGFLIT  
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN  
101 QMRKTVELSA VFLSNIYLG F QQGYFDLSAD ENPVLHIWSL AVEEQYLLY  
151 PLLLIFCCKK TKSLRVLRLNI SIILFLILTA TSFLPSGFYT DILNQPNYY  
201 LSTLRFPELL AGSLLAVYGQ TQNGRRQTAN GKRQLLSSLC FGALLACL FV  
35 251 IDKHNPFIPG MTLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY  
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLI SYYLIEQPLR  
351 KRKMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHLRPL PGAPLAAENH  
401 FPETVLTLD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD  
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGPPARFR  
40 501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG  
551 KSNQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY  
601 YMGREFHKHE RLLKSSRDGA LQ\*

ORF128a (SEQ ID NO: 832) and ORF128-1 (SEQ ID NO: 830) show 99.5% identity in 622 aa  
45 overlap:

orf128a.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNLRWLPGGFLGVDIFFVISGFLITGIILSEIQNG  
|||  
orf128-1 MQAVRYRPEIDGLRAVAVLSVMIFHLNLRWLPGGFLGVDIFFVISGFLITGIILSEIQNG  
|||  
50 orf128a.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG  
|||  
orf128-1 SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG  
|||  
orf128a.pep QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLRLNISIILFLILTA  
|||

	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
	orf128-1	SSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
5	orf128a.pep	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVGKISY
	orf128-1	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFVGKISY
	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
10	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLT LGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLT LGDSHAGHLRGFL
15	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSF LIPGF PPARFRET VKRIA AVKPVYVFANNTSISR SPLREEKLKRFAANQYL
	orf128-1	PVPRFEAQSF LIPGF PPARFRET VKRIA AVKPVYVFANNTSISR SPLREEKLKRFAANQYL
20	orf128a.pep	RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
25	orf128-1	YMGREFHKHERLLKSSHGALQX

### Homology with a predicted ORF from *N. gonorrhoeae*

ORF128 (SEQ ID NO: 828) shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) (SEQ ID NO: 834) from *N. gonorrhoeae*:

30	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
	orf128.pep	LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNISI	90
	orf128ng	LSNIYLG FRLGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCYKKTKSLRVLNRNISI	172
35	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK	232
	orf128.pep	RQLLSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
40	orf128ng	RQLLSLLCFGALLVCLFVIDKHDPFI PGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292



-579-

```

orf128.pep  VFVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA 244
             |||||||
orf128ng    VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR 352

```

5 The complete length ORF128ng nucleotide sequence (SEQ ID NO: 833) is:

```

1  ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC
51  CGTGCATATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101  GATTCCCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC
151  AACATATATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTTT TCCGGGATTT
10  201  TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT
251  CCCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTCAAC
301  CAAATGAGGA AAACCATAGA GCTTCTACG GTTTTTTGT CCAATATTTA
351  TTTGGGGTTC CGATTGGGGT ATTTTCGATT GAGTGCCGAC GAGAACCCCG
401  TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT
15  451  CCTCTTTTGC TGATATTCTG TTACAAAAAA ACCAAATCAC TACGGGTGCT
501  GCGTAATATC AGCATATCC TGTTCCTGAT TTGACCGCA TCATCGTTTT
551  TGCCGCGCGG GTTTTATACC GACATCCTCA ACCAACCCaa TACTTATTAC
601  CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTCGC TGTGGCGGT
651  TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAAACGGC
20  701  AGTTGCTTTC ATTACTCTGT TTCGCGCgCat tgCTTGTCTG CCTGTTTCGTG
751  ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCCTGC TCCTTCCCTG
801  CCTGCTGACG GCGCTGCTTA TCCGAGTAT GCAATACGGG ACATTCCGA
851  CCCGCATCCT GTCGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901  TCCCTATACC TGTACCATTG GATTTTATT GCCTTCGCCC ATTACATTAC
25  951  AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
1001  CGGCCGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA
1051  AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC
1101  GTCCCTGATG CTGTGCGGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
1151  AGGAACCGCT CCGCCGCTG CCCGCGACGC CCGTTGCTGC GGAAAATAAT
30  1201  TTTCCGAAA CCGTCTTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG
1251  GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGAAA GCTAAAATCC
1301  TGTCCCTCGA TTCGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC
1351  AACCCGTTGT GCCGAAAATA CCGGATGAA GTTGAAAAAG CCGAAGCTGT
1401  TTTCAATGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
35  1451  GATTTGAAGC GCAATCCTTC CTGATACCCG GGTTCAAAGC CCGATTCCAG
1501  GAAACCGTCA AGAGGATAGC CGCCGTCAAA CCTGTATATG TTTTGTCAAA
1551  CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT
1601  TTGCTATAAA CCAATACCTC CGGCCTATTC GGGCTATGGG CGACATCGGC
1651  AAGAGCAATC AGGCGTCTT TGATTGGGTT AAAGATATTC CCAATGTGCA
40  1701  TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG
1751  GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801  TATATGGGGC GGAATTTCa CAAACACGAA CGCCTGCTCA AGCATTCCCG
1851  AGGCGGCGCA TTGCACTAG

```

45 This encodes a protein having amino acid sequence (SEQ ID NO: 834):

```

1  MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGGFLGVDI FFVISGFLIT
51  NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101  QMRKTIELST VFLSNLYLGF RLGYFDLSAD ENPVLHIWSL AVEEQYYLLY
151  PLLLIIFYKK TKSLRVLRLNI SIILFLILTA SSFLPAGFYT DILNQPNNTYY
50  201  LSTLRFPPELL VGSLLAVYGQ TQNGRRQTEN GKRQLLSLLC FGALLVCLFV
251  IDKHDPIPG ITLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301  SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
351  KRKMTFFKAF FCLYLAPSLM LVGYNLYSRG ILKQEHLRPL PGPVAAENN
401  FPETVLTLGD SHAGHLRGFL DYVGREGWK AKILSLDSEC LVWVDEKLAD
55  451  NPLCRKYRDE VEKAEVFIA QFYDLRMGGQ PVPREFEQSF LIPGFKARFR
501  ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG
551  KSNQAVFDLV KDIPNVHWVD AQKYLKNTV EIHGRYLYGD QDHLTYFGSY

```

601 YMGREFHKHE RLLKHSRGGALQ\*

ORF128ng (SEQ ID NO: 834) and ORF128-1 (SEQ ID NO: 830) show 95.7% identity in 622 aa overlap:

```

5      orf128-1.pep  MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
      orf128ng      MQAVRYRPEIDGLRAVAVLSVIIIFHLNNRWLPGGFLGVDIFFVISGFLITNIILSEIQNG

      orf128-1.pep  SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
10     orf128ng      SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

      orf128-1.pep  QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLRNISIILFLILTA
      orf128ng      RLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKTKSLRVLRNISIILFLILTA

      orf128-1.pep  SSFLPSGIFYTDILNQPNNTYYLSTLRFPELLAGSLLAVYGQTONGRRQTANGKRQLSSL
15     orf128ng      SSFLPAGIFYTDILNQPNNTYYLSTLRFPELLVGSLLAVYGQTONGRRQTENGKRQLSLLC

      orf128-1.pep  FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
      orf128ng      FGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY

20     orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
      orf128ng      SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF

      orf128-1.pep  FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
25     orf128ng      FCLYLAPSLMLVGYNLYSRGILKQEHLRPLPGTPVAAENHFPETVLTGLGDSHAGHLRGFL

      orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
      orf128ng      DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

      orf128-1.pep  PVPRFEAQSFILIPGFPARFRETVKRIA AVKPVYVFANNTSISRSPREEKLKRFAANQYL
30     orf128ng      PVPRFEAQSFILIPGFKARFRETVKRIA AVKPVYVFANNTSISRSPREEKLKRFAINQYL

      orf128-1.pep  RPIQAMGDIGKSNQAVFDLIKDI PNHVWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
      orf128ng      RPIRAMGDIGKSNQAVFDLVKDI PNHVWVDAQKYL PKNTVEIHGRYLYGDQDHLTYFGSY

35     orf128-1.pep  YMGREFHKHERLLKSSHGGALQX
      orf128ng      YMGREFHKHERLLKHSRGGALQX
                        610      620

```

40 In addition, ORF128ng (SEQ ID NO: 834) shows homology to a hypothetical *H.influenzae* protein (SEQ ID NO: 1164):

sp|P43993|Y392\_HAEIN HYPOTHETICAL PROTEIN HI0392 )gi|1074385|pir||B64007  
hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)

)gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus influenzae] Length = 245

Score = 239 bits (604), Expect = 3e-62

Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

5 Query: 38 VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97  
+DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y  
Sbjct: 1 MDIFFVISGFLITGIIITEIQNSFSLSKQFYTRRIKRIYPAFITVMALVSFIASAIPIYN 60

Query: 98 DFNQMRKTIELSTVFLSNIYLGFRGLGYFDLSADENPVLHIWLSLAVEEQXXXXXXXXXXIFC 157  
DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWLSLAVE Q I  
10 Sbjct: 61 DFNKLRKTIELAIAFLSNFYLGGLTQGYFDLSANENPVLHIWLSLAVEGQYYLIFPLILILA 120

Query: 158 YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPELLVGSLLAV 217  
YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+  
Sbjct: 121 YKKFREVKVLFIIITLILFFILLATSFVSANFYKEVLHQPNIYYLSNLRPELLVGSLLAI 180

Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPGIT 262  
Y N + Q +L++L L CLF+++ + FIPGIT  
15 Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful  
20 antigens for vaccines or diagnostics, or for raising antibodies.

### Example 99

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 835):

25 1 ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT  
51 GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT  
101 TGGCGTTGGC GCGCCTGATT CACTTGGA AAGCCGGTGC GCCGATGCGC  
151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT  
201 CCGGGGTACG CCGCTGTTT TGCAGATTGT GATTGGGCG TATGTGTGTT  
251 TTCCGTTTTT CGTC..

30 This corresponds to the amino acid sequence (SEQ ID NO: 836; ORF129):

1 ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR  
51 VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFV..

Further work revealed the complete nucleotide sequence (SEQ ID NO: 837):

35 1 ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTCCTTTA  
51 CCGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT  
101 CCGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA  
151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT  
201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
40 251 TTTGGGCGTA TGTGTGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT  
301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG

-582-

5  
 451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 501 GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG  
 601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

This corresponds to the amino acid sequence (SEQ ID NO: 838; ORF129-1):

10  
 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI  
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA  
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL  
 15  
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 (SEQ ID NO: 836) shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) (SEQ ID NO: 840) from strain A of *N. meningitidis*:

20  
 orf129.pep                   10           20           30           40           50  
                   IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129a           MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
                   10           20           30           40           50           60

25  
 orf129.pep                   60           70           80  
                   ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFV  
 orf129a           ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
                   70           80           90           100           110           120

30  
 orf129a           SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS  
                   130           140           150           160           170           180

The complete length ORF129a nucleotide sequence (SEQ ID NO: 839) is:

35  
 1 ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTTCTTTA  
 51 CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT  
 101 CGGTATTGGG TCTGTTGTTG GCGTTGCGGC GCCTGATTCA CTTGGAAGAAA  
 151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
 201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA  
 251 TTTGGGCGTA TGTGTGTTT CCGTTTTCG TCCATCCTTC AGACGGCATT  
 301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
 351 GATTGCCGGT TCTTTGGCAC TGATCGCAA CTCGGGGGCG TATATCTGTG  
 401 AGATTTTCCG CGCGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
 451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 45  
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG  
 601 GCGTATGTTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence (SEQ ID NO: 840):

```

1  MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201 AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

ORF129a (SEQ ID NO: 840) and ORF129-1 (SEQ ID NO: 838) show 100.0% identity in 248 aa overlap:

```

orf129a.pep  MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
orf129-1      MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW

orf129a.pep  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGI LVSGEAAIALRRGYGPLIAG
orf129-1      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGI LVSGEAAIALRRGYGPLIAG

orf129a.pep  SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS
orf129-1      SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS

orf129a.pep  EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE
orf129-1      EFITLLKDSS LLSVIAVAEL AYWQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE

orf129a.pep  KRYNPQHRX
orf129-1      KRYNPQHRX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 (SEQ ID NO: 836) shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) (SEQ ID NO: 842) from *N.gonorrhoeae*:

```

orf129.pep      I IYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54
orf129ng        MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60

orf129.pep      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFV 88
orf129ng        ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVILHTAFLGNAMQSRVRPDKGRWIAG 120

```

An ORF129ng nucleotide sequence (SEQ ID NO: 841) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 842):

```

1  MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVILHTAF
101 LGNAMQSR R VPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTEPRN

```

151 PLSMGQRRFP GCENWYPPQN FIKK\*

Further work revealed the following gonococcal sequence (SEQ ID NO: 843):

5	1	ATGGATTTTc	gtTTTGACAT	TATTTAcgaA	TACCGCTGGA	TGTTTCTTTA
	51	CGGCGCACTG	Acgaccttg	ggctgacggt	cgtggcgacg	gCGGGCGGTT
	101	CGGtattggG	TCTGTTGTG	GCGTTGCGC	GCCTGATTCA	CTTGAAAAA
	151	GCCGGTGCGC	CGATGCGCGT	GCTGGCGTGG	GCCTTGCGTA	AGGTTTCGCT
	201	GCTGTACGTT	ACCTGTTCC	GGGGTACGC	GCTGTTTGTG	CAGATTGTGA
10	251	TTTGGGCGTA	TGTGTGGTTT	CCGTTTTTCG	TCCATCCTTC	AGACGGCAT
	301	TTGGTACAGC	GCGAGGCGGC	AATCGCGCTG	CGTCGCGGAT	ACGGGCCGCT
	351	GATTGCCGGT	TCTTTGGCAC	TGATCGCCAA	CTCGGGGGCG	TATATCTGTG
	401	AGATTTTCCG	CGCGGGCATC	CAGTCTATAG	ACAAAGGACA	GATGGAGGCG
	451	GCGTGTCTT	TGGGACTGAC	CTATCCGAG	GCGATGCGCT	ATGTGATTCT
15	501	GCCGCAGGCA	TTGCGCCGTA	TGCTGCCGCC	TTTGCGGAGC	GAGTTCATCA
	551	CGCTCTTGAA	AGACAGCTCG	CTGCTGTGCG	TCATTGCTGT	GGCGGAGTTG
	601	GCGTATGTT	AGAATACAGT	TACGGGCCGG	TATTCGGTTT	ATGAAGAACC
	651	GCTTTACACC	GCCGCCCTGA	TTTATCTGTT	GATGACGACT	TTCTTAGGCT
	701	GGATATTCCT	GCGTTTGGA	AAACGTTACA	ATCCGCAACA	CCGCTGA

20 This corresponds to the amino acid sequence (SEQ ID NO: 844; ORF129ng-1):

25

1	MDFRFDIIYE	YRWMFLYGAL	TTLGLTVVAT	AGGSVLGLLL	ALARLIHLEK
51	AGAPMRVLAW	ALRKVSLLYV	TLFRGTPLFV	QIVWAYVWF	PPFVHPSDGI
101	LVSGEAATAL	RRGYGPLIAG	SLALIANSGA	YICEIFRAGI	QSIDKGQMAE
151	ARSLGLTYPQ	AMRYVILPQA	LRRMLPPLAS	EFITLLKDSS	LLSVIAVAEL
201	AYVQNTITGR	YSVYEEPLYT	VALIYLLMTT	FLGWIFLRLE	KRYNPQHR*

ORF129ng-1 (SEQ ID NO: 844) and ORF129-1 (SEQ ID NO: 838) show 99.2% identity in 248 aa overlap:

```

30      orf129-1.pep  MDFRFDI IYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |
      |
      |
      orf129ng-1    MDFRFDI IYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |
      |
      |
      orf129-1.pep  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAG
      |
      |
      |
      orf129ng-1    ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAG
      |
      |
      |
35      orf129-1.pep  SLALIANS GAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS
      |
      |
      |
      orf129ng-1    SLALIANS GAYICEIFRAGIQSIDKGQMEAAACSLGLTYPQAMRYVILPQALRRMLPPLAS
      |
      |
      |
      orf129-1.pep  EFITLLKDSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
      |
      |
      |
40      orf129ng-1    EFITLLKDSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLE
      |
      |
      |
      orf129-1.pep  KRYNPQHRX
      |
      |
      |
      orf129ng-1    KRYNPQHRX

```

45 In addition, ORF129ng-1 (SEQ ID NO: 844) is homologous to an ABC transporter (SEQ ID NO: 1165) from *A.fulgidus*:

2650409(AE001090) glutamine ABC transporter, permease protein (glnP) [Archaeoglobus fulgidus]Length = 224

Score = 132 bits (329), Expect = 2e-30

Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

5 Query: 65 VSLLYVTFLFRGTPLFVQIVIWVWFPPFVHPSDGILVSGEAAIALRRGYGPLIAGSLAL 124  
+S YV + RGTPL VQI+I +F P+ GI + E A G +AL  
Sbjct: 58 ISTAYVEVIRGTPLLQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99

10 Query: 125 IANSGAYICEIFRAGIQSIDKGQMEAACSLGLTYPQAMRYVILPQALRRMLPPLASEFIT 184  
SGAYI EI RAGI+SI QMEAA SLG+TY QAMRYVI PQA R +LP L +EFI  
Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159

Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLEKR 242  
LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+  
Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

- 15 This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 100

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 845):

20 1 ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA  
51 TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT  
101 GGCTGCCCCG GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC  
151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTT TACGTAAACA  
201 cTACGTCCGC ACTTATTACy TGCTCCAAC TTTGCCGCC GCAGgcTAgT  
25 251 TTGTGGACAG GCGCGGCGwA ATTACAAAC CTGCCCGCyT CCGCGCCCCT  
301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc  
351 TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTGA CTACCCCAA  
401 CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC  
451 TTTCTTGrTG AACGTGAACC CGrTATTTT CATTACCGTT CCTGCGATTC  
30 501 TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATT  
551 CGGGCGAATG CGTTTACAGA CGATCCGGAr Tar

This corresponds to the amino acid sequence (SEQ ID NO: 846; ORF130):

35 1 ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI  
51 LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPASAPL  
101 HLITLGGMMG GVMVWLTA LWHSGFTKLD YPKLCRIAVP ILFAAAVSRA  
151 FLXNVNPXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE\*

Further work revealed the complete nucleotide sequence (SEQ ID NO: 847):

40 1 ATGCGGCCGT TTTTCGTCGG CGCGCGGGTG CTGCCATAC TCGGTGCGCT  
51 GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT  
101 TGGAACCTAT GCTGCCGGCG GCATACGGCG GTTTTTTGAC TGCGGCTTTG  
151 TTGGA CTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCG CGACTTTGAT  
201 GGCGGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAA

-586-

5  
10  
15

```

251 CTGCCTCGTT TTTCGTCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301 GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
351 GTTACTTGCC GCGTTCCTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
401 ATTTGAACCT GTTGC GCGCG CAAGTGCATC TAAATATGGC GGCGGTGATG
451 TTCGTATCCG TCGCGCTCAG TATTCTTTTG GCGCGGAAG CCCTGAAAGA
501 ATGCCGTCTG AAAGACCCCTG TTTTATTCC AAATATCGTT TATAAAAACA
551 TCGCCATTAC TTTCCTGCTC TTGCACGCCG CCGCCGAAC TGGGCTGCCC
601 GCGCAAACCG CCGGTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651 CAAGCTGCGT GAGCTTCACC ATCACGAACT CTTACGTAAA CACTACGTCC
701 GCACTTATTA CCTGCTCCAA CTCTTTGCCG CCGCAGGCTA TTTGTGGACA
751 GGCGCGGCGA AATTACAAAA CCTGCCCCGCC TCCGCGCCCC TGCACCTGAT
801 TACCCTCGGC GGCATGATGG GCGCGGTGAT GATGGTGTGG CTGACCGCCG
851 GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCCAA ACTCTGCCGC
901 ATTGCCGTCC CCATCTTTT CGCCGCCGCC GTCTCGCGCG CTTTCTTGAT
951 GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
1001 CCGTATTCGT ACTGTATCTT TTCACGTTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence (SEQ ID NO: 848; ORF130-1):

20  
25

```

1 MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLFLFC
101 ARLIWLD RNT DN FALLMLLA APTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAELWLP
201 AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAGYLWT
251 GAAKLQNLPA SAPLHLITLG GMMGGVMVW LTAGLWHS GF TKLDYPKLCR
301 IAVPILFAAA VSRAFLMNVN PIFFITVPAI LTAAVFVLYL FTFIPIFRAN
351 AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

### 30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 (SEQ ID NO: 846) shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) (SEQ ID NO: 850) from strain A of *N. meningitidis*:

35  
40  
45

```

          10      20      30
orf130.pep      LKECRLKDPVFIPNIVYKNIAITFLLHAA
                |||||:|||||
orf130a      LNLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNVVYKNIAITFLLHAA
          140      150      160      170      180      190

          40      50      60      70      80      90
orf130.pep      AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX
                |||||:|||||
orf130a      AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAK
          200      210      220      230      240      250

          100      110      120      130      140      150
orf130.pep      LQNLPA S APLHLITLGGMMGGVMVWLTAGLWHS GF TKLDYPKLCRIAVPILFAAAVSRA
                |||||:|||||
orf130a      LQNLPA S APLHLITLGGMMGSVMVWLTAGLWHS GF TKLDYPKLCRIAVPILFAAAVSRA
          260      270      280      290      300      310

```



```

              160      170      180      190
orf130.pep    FLXNVNPXFFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPEX
              |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf130a       VLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPEX
              320      330      340      350

```

The complete length ORF130a nucleotide sequence (SEQ ID NO: 849) is:

```

      1  ATGCGGCCGT  TTTTCGTCGG  CGCGGCGGTG  CTTGCCATAC  TCGGTGCGCT
     51  GGTGTTTTTC  ATCAACCCCG  GTGCCATCGT  CCTGCACCGC  CAAATTTTCT
    101  TGGAACCTAT  GCTGCCGCGC  GCATACGGCG  GTTTTTTGAC  TGCGGCTTTG
    151  TTGGA CTGGA  CGGGTTTTTC  GGGTAACCTG  AAACCTGTCT  CGACTTTGAT
    201  GGCGGCATTA  TTGCTCGCGC  CATCCGCTAT  ACTGCCCTTT  TCGCCGCAAA
    251  CTGCCTCGTT  TTTCTGTCGC  GCCTATTGGC  TGGTGTGCT  GCTGTTCTGC
    301  GCCCGGCTGA  TTTGGCTAGA  CCGAAACACC  GACAACTTCG  CCCTGCTAAT
    351  GTTACTTGCC  GCGTTCACCT  TTTTTCAGAC  GGCATATGCC  GTCAGCGGCG
    401  ATTTGAACCT  GTTGCGCGCG  CAAGTGCATC  TAAATATGGC  GGCGGTGATG
    451  TTCGTATCCG  TCGCGCTCAG  TATTCTTTTG  GGCGCGGAAG  CCCTGAAAGA
    501  ATGCCGTCTG  AAAGACCCAG  TATTCATCCC  CAATGTCGTC  TATAAAACA
    551  TCGCCATTAC  CTTCTGCTC  CTGCACGCGC  CCGCCGAACT  TTGGCTGCCT
    601  GCGCAAACCG  CCGGTTTTAC  CTCGCTCGCC  GTCGGCTTTA  TCCTGCTTGC
    651  CAAGCTGCGT  GAGCTTCACC  ATCACGAACT  CCTGCGCAAA  CACTACGTCC
    701  GCACTTATTA  CCTGCTCCAA  CTCTTTGCCG  CCGCAGGCTA  TTTGTGGACA
    751  GGCGCGGCGA  AATTACAAAA  CCTGCCCGCC  TCCGCGCCCC  TGCACCTGAT
    801  TACCCTCGGT  GGCATGATGG  GCAGCGTGAT  GATGGTGTGG  CTGACTGCCG
    851  GACTGTGGCA  CAGCGGCTTT  ACCAAGCTCG  ACTACCGGAA  ACTCTGCCGC
    901  ATCGCCGTCC  CCATCCTNTT  CGCCGCCGCC  GTTTCGCGCG  CTGTTTAAAT
    951  GAACGTAAAC  CCGATATTCT  TCATCACCGT  CCCCACAATT  CTGACCGCCG
   1001  CCGTGTTCGT  GCTTTACCTG  CTGACATTCG  TACCGATCTT  TCGGGCGAAC
   1051  GCGTTTACAG  ACGATCCGGA  ATAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 850):

```

      1  MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
     51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVL LFC
    101  ARLIWLD RNT DNFALLMLLA AFTVFQ TAYA VSGDLNLLRA QVHLNMAAVM
    151  FVSVRVS ILL GAEALKECRL KDPVFIPNVV YKNIAITFLL LHAAAE LWLP
    201  AQTAGFTSLA VGFILLAKLR ELHHHELLRK HYVRTY YLLQ LFAAAGYLWT
    251  GAAKLQNLPA SAPLHLITLG GMMGSVMVWV LTAGLWHS GF TKLDYPKLCR
    301  IAVPILFAAA VSRVLMNVN PIFFITVPAI LTAAVFVLYL LTFVPIFRAN
    351  AFTDDPE*

```

ORF130a (SEQ ID NO: 850) and ORF130-1 (SEQ ID NO: 848) show 98.3% identity in 357 aa overlap:

```

orf130a.pep    MRPFFVGA AVLAILGALVFFINPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL
              |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf130-1       MRPFFVGA AVLAILGALVFFINPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL

orf130a.pep    KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFALLMLLA
              |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf130-1       KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFALLMLLA

orf130a.pep    AFTVFQ TAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNVV
              |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf130-1       AFTVFQ TAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNVV

```

```

orf130a.pep  YKNIAITFLLLHAAELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
              |||
orf130-1     YKNIAITFLLLHAAELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ

orf130a.pep  LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCR
              |||
orf130-1     LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCR

orf130a.pep  IAVPILFAAAVSRVLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPE
              |||
orf130-1     IAVPILFAAAVSRVLMNVNPIFFITVPAILTAAVFVLYLFTFPIFRANAFTDDPE

```

# 10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 (SEQ ID NO: 846) shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) (SEQ ID NO: 852) from *N.gonorrhoeae*:

```

orf130.pep                                LKECRLKDPVFIPNIVYKNIAITFLLLHAA  30
              |||
15 orf130ng  LNLLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNIVYKNIAIT-LLLHAA  201

orf130.pep  AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX  90
              |||
orf130ng    AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX  261

orf130.pep  LQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSFGFTKLDYPKLCRIAVPILFAAAVSR  150
              |||
20 orf130ng  LQNLPAAPLHLITLGGMTGGVMMVWLTAGLWHSFGFTKLDYPKLCRIAVSILFASAVSR  321

orf130.pep  FLXNVNPXFFITVPAILTAAVFVLYLFXFPIFRANAFTDDPE  193
              |||
25 orf130ng  VLMNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPE  364

```

An ORF130ng nucleotide sequence (SEQ ID NO: 851) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 852):

```

1  MNKFFTHPMR PFFVGA AVLA ILGALVFFHQ PRRYHPAPPN FLGTYAAGCI
30 51  RRFDDYRFVG PDGFFRQPET CRYFDGGVVA CCGCFIAVET ATCRIFRRRL
101 LAGVA AVLRL ADLARRQHRT LRSVDVTAAF TVFQTAYAVS GDLNLLRAQV
151 HLNMAAVMFV SVRVSVLLGT ETLKECRLKD PVFIPNVIYK NIAITLLLHA
201 AAELWLPAQT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYYLLQLFA
251 AAGYLWTGAA KLQNLPAAP LHLITLGGMT GGVMVWLTG GLWHSFGFTKL
301 DYPKLCRIAV SILFASAVSR AVLNMVNPIF FITVPEILTA AVFMLYLLTF
35 351 VPIFRANAFT DDPE*

```

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 853):

```

1  ATGCGCCCGT TTTTCGTCGG TCGGCAGTA CTTGCCATAC TCGGTGCGTT
40 51  GGTGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
101 TGGA ACTTAT GCTGCCGGCT GCATACGGCG GTTTTTTGAC TACCGCTTTG
151 TTGGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
201 GCGCGTGTG TTGCTTGTTG CGGCTGTTT ATTGCCGTTT TTACCGCAAC
251 TTGCCGCATT TTCGTCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC

```

5  
10  
15

```

301 GCCTGGCTGA TTTGGCTCGA CCGCAACACC GACAACTTCG CTCTGTTGAT
351 GTTACTTGCC GCATTACCG TTTTTCAGAC GGCTATGCC GTCAGCGGCG
401 ATTTGAACTT ACTGCGCGCG CAAGTGCATT TGAATATGGC GGCGGTCATG
451 TTCGTATCCG TCCGCGTCAG CGTCCTTTTG GGCACGGAAC CCCTGAAAGA
501 ATGCCGTCTG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAACA
551 TCGCCATCAC CCTGCTGCTG CACGCCGCGG CCGAACTTTG GCTGCCCGCG
601 CAAACCGCCG GTTTTACTGC GCTTGCCGTC GGCTTCATCC TGCTCGCCAA
651 GCTGCGCGAA CTGCACCATC ACGAACTCTT ACGCAAACAC TACGTCCGCA
701 CTTATTACCT GCTCCAGCTC TTTGCGCGCG CAGGTTATCT GTGGACAGGC
751 GCGGCGAAAC TGCAAAACCT GCCCGCCTCC GCGCCCCTGC ACCTGATTAC
801 CCTCGGCGGC ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCCGGAC
851 TGTGGCACAG CGGCTTACC AAACCTGACT ACCCGAACT CTGCCGCATC
901 GCCGTCTCCA TCCTTTTCGC CTCCGCCGTT TCGCGCGCTG TTTTAATGAA
951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG
1001 TGTTTCATGCT TTACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCG
1051 TTTACAGACG ATCCGGAATA A

```

This corresponds to the amino acid sequence (SEQ ID NO: 854; ORF130ng-1):

20  
25

```

1 MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELMLPA AYGGFLTTAL
51 LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVL LLLFC
101 AWLIWLD RNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAELWLPA
201 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYYLLQL FAAAGYLWTG
251 AAKLQNL PAS APLHLITLGG MTGGVMVWL TAGLWHS GFT KLDYPKLCRI
301 AVSILFAS AV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA
351 FTDDPE*

```

ORF130ng-1 (SEQ ID NO: 854) and ORF130-1 (SEQ ID NO: 848) show 92.4% identity in 357 aa overlap:

30  
35  
40  
45

```

orf130-1.pep MRPFFVGA AV LAILGALVFF INPGAIV LHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL
orf130ng-1 MRPFFVGA AV LAILGALVFF INPGAII LHRQIFLELMLPAAYGGFLTTALLDRTGFSGNL

orf130-1.pep KPVATLMAALLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLD RNTDNFALLMLLA
orf130ng-1 KPAATLMAVLLLVAAVLLPFLPQLAAFFVAAYWLVL LLLFCAWLIWLD RNTDNFALLMLLA

orf130-1.pep AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNIV
orf130ng-1 AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVS VLLGTETLKECRLKDPVFIPNVI

orf130-1.pep YKNIAITFLLLHAAAE LWLPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYYLLQ
orf130ng-1 YKNIAIT-LLLHAAAE LWLPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYYLLQ

orf130-1.pep LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMMGGVMVWL TAGLWHS GFTKLDYPKLCR
orf130ng-1 LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMVWL TAGLWHS GFTKLDYPKLCR

orf130-1.pep IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLFTFPIFRANAFTDDPEX
orf130ng-1 IAVSILFASAVSRAVLMNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 101

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 855):

```

5          1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
          51  TACGTTGCA  GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
        101  CCGGCTGGTG  TAAGCCGAGA AAACCGGCTG CCATCGATT  TTGGGATATT
        151  GGC GCGGAGA GTCCGCCGTC TTAGGGGAC  TACGAGATAC CGCTTTCAGA
        201  CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
10        251  ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
        301  ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
        351  CTGCTTGAA  AAG. .

```

This corresponds to the amino acid sequence (SEQ ID NO: 856; ORF131):

```

15          1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
          51  GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSIFYRKIG KFEXCGLDWR
        101  TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence (SEQ ID NO: 857):

```

20          1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
          51  TACGTTGCA  GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
        101  CCGGCTGGTG  TAAGCCGAGA AAACCGGCTG CCATCGATT  TTGGGATATT
        151  GGC GCGGAGA GTCCGCCGTC TTAGGGGAC  TACGAGATAC CGCTTTCAGA
        201  CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
25        251  ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
        301  ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
        351  CTGCTTGAA  AAGCAGGGGT TGC GCGCAA  CGGTCTGTCC GAGCGCGTCC
        401  GATGGTAA

```

30 This corresponds to the amino acid sequence (SEQ ID NO: 858; ORF131-1):

```

          1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
          51  GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSIFYRKIG KFEACGLDWR
        101  TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*

```

35 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 (SEQ ID NO: 856) shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) (SEQ ID NO: 860) from strain A of *N. meningitidis*:

-591-

```

      10      20      30      40      50      60
orf131.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
      |||||
orf131a    MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLED
      10      20      30      40      50      60

      70      80      90      100     110     120
orf131.pep YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL
      |||||
orf131a    YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
      70      80      90      100     110     120

orf131.pep K
      |
orf131a    KQGLRRNGLSERVRWX
      130

```

The complete length ORF131a nucleotide sequence (SEQ ID NO: 859) is:

```

      1  ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
     51  TACGGTTGCA GGCTGCCGGT TGCAGGTTG GTATGAGTGT TCGTCCCTGT
    101  CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTTGGGATATT
    151  GCGCGCGAGA GTCCTCCGTC TTAGAGGAC TACGAGATAC CGCTTTCAGA
    201  CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
    251  ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
    301  ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTTGA
    351  TTGTTTGAAG AAGCAGGGGT TCGGCGCAA CGGTCTGTCC GAGCGCGTCC
    401  GATGGTAA

```

This encodes a protein having amino acid sequence (SEQ ID NO: 860):

```

      1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
     51  GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
    101  TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW*

```

ORF131a (SEQ ID NO: 860) and ORF131-1 (SEQ ID NO: 858) show 97.0% identity in 135 aa overlap:

```

    35  orf131a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLED
      |||||
      orf131-1    MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

      orf131a.pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
      |||||
    40  orf131-1    YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL

      orf131a.pep KQGLRRNGLSERVRWX
      |||||
      orf131-1    KQGLRRNGLSERVRWX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 (SEQ ID NO: 856) shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) (SEQ ID NO: 862) from *N.gonorrhoeae*:

```

5      orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
      orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED  60

      orf131.pep  YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTDVGKPLIETFKQGGFDCLE  120
      orf131ng    YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTDVGKPLVERFKQEGFDCLE  120

10     orf131.pep  K 121
      orf131ng    KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence (SEQ ID NO: 861) was predicted to encode a protein having amino acid sequence (SEQ ID NO: 862):

```

15      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
      51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
      101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 863):

```

20      1 ATGGAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTTGCAAT
      51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
      101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
      151 GCGGCGGAGA GtccgctGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
      201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
25     251 ACTTTTATAG GAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
      301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
      351 CTGTTTGGA AAGCAGGGGT TCGGCGCAA CGGCCTGTCC GAGCGCGTCC
      401 GATGGTAA

```

30 This corresponds to the amino acid sequence (SEQ ID NO: 864; ORF131ng-1):

```

      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
      51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
      101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

35 ORF131ng-1 (SEQ ID NO: 864) and ORF131-1 (SEQ ID NO: 858) show 92.6% identity in 135 aa overlap:

```

40     orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED
      orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

      orf131ng-1.pep YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTDVGKPLVERFKQEGFDCLE
      orf131-1      YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTDVGKPLIETFKQGGFDCLE

```

```

orf131ng-1.pep KQGLRRNGLSERVRWX
               |||||
orf131-1       KQGLRRNGLSERVRWX

```

- 5 Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 102

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 865)

```

10      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
      51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101    101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
      151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
      201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
15      251  TGAACCTCGG CCTGCCtTAT AtTtCGGCC CGCAATGGCT GTCGGAAAAC
      301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
      351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCTTCGCGC
      401  CGGGCTTCCT TAtGGCGGC GTACC.GGAA AATtCGGCG TTTCCGCCCG
45      451  CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
      501  TCATCGAAGC CGACGAATAC GACACCGCCT TtTCGACAA AC GTTCTAAA
      551  TtCGTGCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAATTCTGA
      601  CCAGCCGAC ATCTTTGCCG ACTGGGCGC GATACAGACc CAGTTCCACT
      651  ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
      701  CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGa
25      751  AAAATTCTGGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence (SEQ ID NO: 866; ORF132):

```

30      1  MKHIHIIGIG GTFMGGLAAl AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
      51  YEGFDAAQLD EFKADVYVIG NVAKRGMDV V EAILNLGLPY ISGPQWLSEN
101    101  VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
      151  PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
      201  PRRHLRLGR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
      251  KIRHGTRLA..

```

- 35 Further work revealed the complete nucleotide sequence (SEQ ID NO: 867):

```

40      1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
      51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101    101  AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
      151  TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
      201  CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
      251  TGAACCTCGG CCTGCCtTAT ATTTCCGGCC CGCAATGGCT GTCGGAAAAC
      301  GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
      351  GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCTTCGCGC
45      401  CGGGCTTCCT TATGGCGGC GTACCGGAAA ATTCGGCGT TTTCCGCCCGC
      451  CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
      501  CATCGAAGCC GACGAATACG ACACCGCCTT TTTGACAAA CGTTCTAAAT
      551  TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
      601  CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCCACTA

```

```

5   651 CCTCGTGCCT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
    701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
    751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
    801 CTCGTTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
    851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
    901 GCGCGTCATG TCGGTGTCGA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
    951 GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA
10  1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
    1051 ATTCAAGGTT TCGGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
    1101 CGAACC CGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
    1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGGCGTG
    1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
    1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAG AACGCCGAAG
    1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
15  1351 GGAAAGCTGC TGAAGCTTT GAGATAG

```

This corresponds to the amino acid sequence (SEQ ID NO: 868; ORF132-1):

```

20  1  MKHIHIIGIG GTFMGGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
    51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSEN
    101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
    151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
    201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQSLQDT LDKGCWTPVE
    251 KFGTEHWQQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
    301 ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPATAIETT
25  351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLEKA DQVFCYAGGV
    401 DWDVAEALAP LGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH
    451 GKLEALR*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with the hypothetical o457 protein (SEQ ID NO: 1166) of *E.coli* (accession number U14003)

ORF132 (SEQ ID NO: 866) and o457 (SEQ ID NO: 1166) show 58% aa identity in 140 aa overlap:

```

35  Orf132: 4  IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLDEFK 63
    IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
    o457: 3  IHILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61

    Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTASML 123
    D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M
    o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121

    Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143
    W+LE G PGF+IGGV G
40  o457: 122 TWILEQCGYKPGFVIGGVPG 141

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF132 (SEQ ID NO: 866) shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) (SEQ ID NO: 870) from strain A of *N. meningitidis*:



		10	20	30	40	50	60
	orf132.pep	MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGDAKMYPPMSTQLEALGIDVYEGFDAQLD					
5	orf132a	MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGDAKMYPPMSTQLEALGIGVYEGFDTAQLD					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA					
10	orf132a	EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWLVGVAXTHGKTTTA					
		70	80	90	100	110	120
		130	140	150	160		
	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR					
15	orf132a	SMLAWVLEYAGLAPGFXIGGVPENFSVSARL-PQTPRQDPNSQSPFFVIEADEYDTAFFD					
		130	140	150	160	170	
		170	180	190	200	210	220
	orf132.pep	HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCLGRDTPVPLPRAYRAVXRL					
20	orf132a	KRSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQLQD					
		180	190	200	210	220	230

The complete length ORF132a nucleotide sequence (SEQ ID NO: 869) is:

	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGTGGGAT
25	51	TGCCGCCATT	GCCAAAGAAG	CAGGGTTTGA	ANTCAGCGGT	TGCGATGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTG
	151	TATGAAGGCT	TCGACACCGC	GCAGTTGGAC	GAATTTAAAG	CCGACGTTTA
	201	CGTTATCGGC	AATGTCGCCA	AGCGCGGGAT	GGATGTGGTT	GAAGCGATTT
	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAAC
30	301	NTGCTGCACC	ATCATTGGNN	ACTCGGCGTG	GCGGNGACGC	ACGGCAAAAC
	351	GACCACCGCG	TCTATGCTCG	CGTGGGTTTT	GGAATATGCC	GGACTCGCAC
	401	CGGGCTTCNT	TATCGGCGGC	GTACCGGAAA	ACTTCAGCGT	TTCCGCCCGC
	451	CTGCCGCAAA	CGCCGCGCCA	AGACCCGAAC	AGCCAATCGC	CGTTTTTCGT
	501	CATTGAAGCC	GACGAATACG	ACACCGCGTT	TTTCGACAAA	CGTCCAAAT
35	551	TCGTGCATTA	CCGTCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	TTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGCCT	ACCGTGCCGT	CTGAAGGCCT	CATCGTCTGC	AACGGACGGC
	701	AGCAAAGCCT	GCAAGACACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGA
	751	AAATTCGGCA	CGGAACACGG	CTGGCAGGCC	GGCGAAGCCA	ATGCCGATGG
40	801	CTCGTTCGAC	GTGTTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCGCTTGGA
	851	GTTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCNGT	CATCGCCGCC
	901	GCGCGTCATG	CCGGAGTNGA	CATTGAGACG	GCCTGCGAAG	CCTTGAGCAC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAAT	CAAAGGCACG	GCAAACGGTA
	1001	TCACCGTTTA	CGACGACTTC	GCCCACCATC	CGACCGCTAT	CGAAACCACG
45	1051	ATTCAAGGTT	TGCGCCAGCG	CGTCGGCGGC	GCGCGCATCC	TCGCCGTCCT
	1101	CGAACCGCGT	TCCAATACGA	TGAAGCTGGG	TACGATGAAA	GCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGNTACGC	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCGGCA	GGCTGCACGT
	1251	CGGCAAAGAC	TTCGATTGCT	TCGTTGCCGA	AATCGTGAAA	AACGCCGAAG
50	1301	CAGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACCTGC	TGGACGCTTT	GAGATAG		

This encodes a protein having amino acid sequence (SEQ ID NO: 870):

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV

51 YEGFDTAQLD EFKADVYVIG NVAKRGMDDV EAILNRGLPY ISGPQWLAEN  
 101 XLHHHWXLG V AXTHGKTTTA SMLAWVLEYA GLAPGFXIGG VPENFVSAR  
 151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD  
 201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQQLQDT LDKGCWTPVE  
 251 KFGTEHGWQA GEANADGSFD VLLDGKKAGH VAWSLMGGHN RMNALAVIAA  
 301 ARHAGVDIQT ACEALSTFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT  
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK AALPASLKEA DQVFXYAGGA  
 401 DWDVAEALAP LGGRLHVGKD FDAFVAEIVK NAEAGDHILV MSNGGFGGIH  
 451 TKLLDALR\*

ORF132a (SEQ ID NO: 870) and ORF132-1 (SEQ ID NO: 868) show 93.9% identity in 458 aa overlap:

orf132a.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAMYPMPSTQLEALGIGVYEGFDTAQLD  
 orf132-1 MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD  
 orf132a.pep EFKADVYVIGNVAKRGMDDVEAILNRGLPYISGPQWLAENXLHHHWXLG VAXTHGKTTTA  
 orf132-1 EFKADVYVIGNVAKRGMDDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA  
 orf132a.pep SMLAWVLEYAGLAPGFXIGGVPENFVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK  
 orf132-1 SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK  
 orf132a.pep RSKFVHYRPR TAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQLQDT  
 orf132-1 RSKFVHYRPR TAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQLQDT  
 orf132a.pep LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA  
 orf132-1 LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNMNALAVIAA  
 orf132a.pep ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG  
 orf132-1 ARHVGVDIQTACEALGAFKNVRRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG  
 orf132a.pep ARILAVLEPR SNTMKLGTMK AALPASLKEADQVFXYAGGADWDVAEALAPLGRLHVGKD  
 orf132-1 ARILAVLEPR SNTMKLGTMK SALPVSLKEADQVFCYAGGVDWDVAEALAPLGRLNVGKD  
 orf132a.pep FDAFVAEIVK NAEAGDHILVMSNGGFGGIHTKLLDALRX  
 orf132-1 FDAFVAEIVK NAEVGDHILVMSNGGFGGIHGKLLLEALRX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 (SEQ ID NO: 866) shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) (SEQ ID NO: 872) from *N. gonorrhoeae*:

40 orf132.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD 60  
 orf132ng MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAMYPMPSTQLEALGIGVHEGFDAAQLE 60

	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHRLRLGRDTPVPPRAHRTIRRPHRLQRTAAKPARY	240
10	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
	orf132ng	FGQRLLDAGGKIRHRTLADW	261

An ORF132ng nucleotide sequence (SEQ ID NO: 871) was predicted to encode a protein having  
15 amino acid sequence (SEQ ID NO: 872):

```

1  MKHIHIIGIG  GTFMGGIAAI  AKEAGFKVSG  CDAKMYPPMS  TQLEALGIGV
51  HEGFDAAQLE  EFQADIYVIG  NVARRGMDV  EAILNRGLPY  ISGPQWLAEN
101 VLHHHWLVGV  AGTHGKTTTA  SMLAWVLEYA  GLAPGFLIGG  VPGKFRFRFP
151 PTANAASRPE  QQIAVFRHRS  RRIHRLFRQ  TLQIRALSPA  YRRVEQSGIR
201 PRRHLRLGR  DTDVPPPPRA  HRTIRRPRL  QRTAAKPARY  FGQRLLDAGG
251 KIRHRTLAD  W*
```

Further work revealed the following gonococcal DNA sequence (SEQ ID NO: 873):

```

25  1  ATGAAACACA  TCCACATTAT  CGGTATCGGC  GGCACGTTTA  TGGGCGGGAT
51  TGCCGCCATT  GCCAAGAAG  CCGGGTTCAG  AGTCAGCGGT  TGCGACGCCA
101 AGATGTATCC  GCCGATGAGC  ACCCAGCTCG  AAGCCTTGGG  CATAGGCCGA
151 CACGAAGGCT  TCGATGCCGC  GCAGTTGGAA  GAATTTCAAG  CCGATATTTA
201 CGTCATCGGC  AATGTCGCCA  GGCGCGGGAT  GGATGTGGTC  GAGGCGATTT
30  251 TGAACCGTGG  GCTGCCTTAT  ATTTCCGGCC  CGCAATGGCT  GGCTGAAAc
301 GTGctgcacc  atcaTTGGgt  ACTCGGCGTG  GcagggacGC  ACGGcaaAac
351 gaccaCcGcg  tCCATGCTCG  CCTGGGTCTT  GGAATATGCC  GGAATCGCGC
401 CGGGCTTCCT  CATCGGCGGt  gtaccggaAA  ATTTCCGGCT  TTCCGCCCGC
451 CTACCGCAAA  CGCCGCGTCA  AGACCCGAAC  AGCAAATCGC  CGTTTTTCGT
501 CATCGAAGCC  GACGAATACG  ACACCGCCTT  TTTTCGACAA  CGCTCCAAAT
35  551 TCGTGCATTA  TCGCCCGCGT  ACCGCCGTGT  TGAACAATCT  GGAATTCGAC
601 CACGCCGACA  TCTTCGCCGA  CTTGGGCGCG  ATACAGACCC  AGTTCCACCA
651 CCTCGTGCGC  ACCGTACCAT  CCGAAGGCCT  CATCGTCTGC  AACGGACAGC
701 AGCAAAGCCT  GCAAGATACT  TTGGACAAAG  GCTGCTGGAC  GCCGGTGGAA
751 AAATTCCGCA  CCGGACACGG  CTGGCAGATT  GGTGAAGTCA  ATGCCGACGG
40  801 CTCGTTTCGAC  GTATTGCTTG  ACGGCAAAAA  AGCCGACAC  GTCGCATGGG
851 ATTTGATGGG  CGGACACAAC  CGCATGAACG  CGCTCGCCGT  CATCGCTGCC
901 GCACGCCATG  CCGGAGTCGA  TGTTAGACG  GCCTGCGAAG  CCTTGGGTGC
951 GTTTAAAAAC  GTCAAACGCC  GCATGGAAAT  CAAAGGCACG  GCAAACGGCA
45  1001 TCACCGTTTA  CGACGATTTC  GCCCACCACC  CGACCGCCAT  CGAAACCACG
1051 ATTCAAGGTT  TGCGCCAACG  TGTCGGCGGC  GCGCGCATCC  TCGCCGTCTT
1101 CGAGCCGCGT  TCCAACACCA  TGAAACTCGG  CACGATGAAG  TCCGCCCTGC
1151 CCGCAAGCCT  CAAAGAAGCC  GACCAAGTGT  TCTGTACGC  CGGCGGCGCG
1201 GACTGGGACG  TTGCCGAAGC  CCTCGCGCCT  TTGGGCTGCA  GGCTGCGCGT
1251 CGGTAAAGAT  TTCGATACCT  TCGTTGCCGA  AATTGTGAAA  AACGCCCCGAA
50  1301 CCGGCGACCA  TATTTTGGTG  ATGAGCAACG  GCGGTTTCGG  CGGAATACAC
1351 ACCAAACTGC  TGGACGCTTT  GAGATAG
```

This corresponds to the amino acid sequence (SEQ ID NO: 874; ORF132ng-1):

```

      1 MKHIHIIGIG GTFMGGIAAI AKEAGFKVSG CDAKMYPPMS TQLEALGIGV
    51 HEGFDAAQLE EFQADIYVIG NVARRGMDVV EAILNRGLPY ISGPQWLAEN
   101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
   151 LPQTPRQDPN SKSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
   201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGQQQSLQDT LDKGCWTPVE
   251 KFGTGHGWQI GEVNADGSFD VLLDGKKAGH VAWDLMGGHN RMNALAVIAA
   301 ARHAGVDVQT ACEALGAFKN VKRMEIKGT ANGITVYDDF AHHPTAIETT
   351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPASLKEA DQVFCYAGGA
   401 DWDVAEALAP LGCRLRVGKD FDTFVAEIVK NARTGDHILV MSNGGFGGIH
   451 TKLLDALR*

```

ORF132ng-1 (SEQ ID NO: 874) and ORF132-1 (SEQ ID NO: 868) show 93.2% identity in 458 aa overlap:

```

15 orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAQLE
   orf132-1      MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD

   orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA
20 orf132-1      EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA

   orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1      SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK

   orf132ng-1.pep RSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGQQQSLQDT
25 orf132-1      RSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQSLQDT

   orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAVDLMGGHNRNALAVIAA
   orf132-1      LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNALAVIAA

   orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTTIQGLRQRVGG
30 orf132-1      ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITVYDDFAHHPTAIETTTIQGLRQRVGG

   orf132ng-1.pep ARILAVLEPRSNTMKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRLRVGKD
   orf132-1      ARILAVLEPRSNTMKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD

   orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
35 orf132-1      FDAFVAEIVKNAEVDGHDILVMSNGGFGGIHKKLLEALRX

```

40 In addition, ORF132ng-1 (SEQ ID NO: 874) is homologous to a hypothetical *E.coli* protein (SEQ ID NO: 1166):

```

pir||S56459 hypothetical protein o457 - Escherichia coli )gi|537075 (U14003)
ORF_o457 [Escherichia coli] )gi|1790680 (AE000494) hypothetical 48.5 kD protein in
fbp-pmba intergenic region [Escherichia coli] Length = 457

```

Score = 474 bits (1207), Expect = e-133  
Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

5 Query: 22 KEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQLEEFQADIYVIGNVARRGMDVVE 81  
++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE  
Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

Query: 82 AILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141  
A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV  
Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCQGYKPGFVIGGV 139

10 Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPTAVLNNLEFDH 201  
P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH  
Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

Query: 202 ADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDQLDKGCWTPVEKFGTGHGWQIG 261  
ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ  
Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWQAK 250

15 Query: 262 EVNADGS-FDVLLDGKKAGHVAWDLMGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320  
++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N  
Sbjct: 251 KLTTDASEWEVLLDGEKVGKWSLVGEHNMHNGLMAIAAARHVGVPADAANALGSFIN 310

20 Query: 321 VKRRMEIKGTANGITVYDDFAHHPTAIETTTIQGLRQRVGG-ARILAVLEPRSNTMKLGTM 379  
+RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNTMK+G  
Sbjct: 311 ARRRLELRGEANGVTYVYDDFAHHPTAILATLAALRGKVGGTARIIAVLEPRSNTMKMGIC 370

Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDHI 438  
K L SL AD+VF W VAE D DT +VK A+ GDHI  
Sbjct: 371 KDDLAPSLGRADEVFLQPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430

25 Query: 439 LVMSNGGFGGIHTKLLDAL 457  
LVMSNGGFGGIH KLLD L  
Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 ORF132-1 (SEQ ID NO: 868) (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein, was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA  
35 (positive result). These experiments confirm that ORF132 (SEQ ID NO: 866) is a surface-exposed protein, and that it is a useful immunogen.

### Example 103

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 875)

1 ..CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA  
 51 CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC  
 101 CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC  
 151 ATTAGTGCGG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTTCGG  
 5 201 CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTC CAAATCGGCG  
 251 ACTCCGGCGT TCACACCGCC TTAAAACCAG AGCGCGCAAA CACTTGGCAA  
 301 TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAAAACAAG ATGATACATT  
 351 AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA  
 401 ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC  
 10 451 AGCAGCACCG GGCTTGCCCTA CACCATCCAA CATCGCrATT TCaWAGACAA  
 501 AGTGCATCAA nnnnnnnnnnn nnnnnnnnnnn nnnnnTACGAT TATGGGCGTT  
 551 TTTTCACCAA CCTTCTTAC GCCTATCAA AAAGCACGCA ACCGACCAAC  
 601 TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT  
 651 CAAACAAGGT TATGGGTGTA GCAGGGTTTC CGCCCTGCCG CGAGATTACG  
 15 701 GACGTTTGA AGTCGGTACG CGCTGGTTGG GCAACAACT GACTTTGGGC  
 751 GGCGCGATGC GCTATTTTCG CAAGAGCATC CGCGCGACGG CTGAAGAACG  
 801 CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTC CGGCAACTGG  
 851 GCAAGCGTTC CATCAAACA ACCGAACTC TTGCCCGCCA GCCTTTGATT  
 901 TTWgATTTTa ACGCGCTTA CGAGCCGAAG AAAACCTTA TTTCCGCGC  
 20 951 CGAAGTCAAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATGCGG  
 1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GTCGTTCGA CCCGAAAGAC  
 1051 AAGGACrrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA  
 1101 ATACGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA  
 1151 CCTTTTgAT GACGATGAGC TACAAGTTTT AA

This corresponds to the amino acid sequence (SEQ ID NO: 876; ORF133):

1 ..PGYYGSDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS  
 51 ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ  
 101 FGFTYKKGL LKQDDTLGLK LVGYRSRIDN YIHNVYKWW DLNGDIPSWV  
 30 151 SSTGLAYTIQ HRFXDKVHQ XXXXXXXXYD YGRFFTNLSY AYQKSTQPTN  
 201 FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTLG  
 251 GAMRYFGKSI RATAEERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI  
 301 XDFNAAYEPK KNLFRAEVK NLFDRRYIDP LDAGNDAAXE RYSSFDPKD  
 351 KDXDVTENAD KTLNKGKYG TSKSVLTNFA RGRTFLMTMS YKF\*

Further work revealed the further partial DNA sequence (SEQ ID NO: 877):

1 GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT  
 51 ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC  
 101 AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC  
 40 151 CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA  
 201 TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA  
 251 TCACGCAGAC CTTTATTTCG ACTTCTACCG ATGCGGGCAG GGCAGGCGGT  
 301 TCATCTCAAT TCGGTGCATC TGTCGACAGC AATTTTATTG CCGGACTGGA  
 351 TGTCGTCAA GGCAGTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG  
 45 401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGCAAT  
 451 AATACCTACG GCCTGCTGCT AAAAGGTCTG ACCGGCACCA ATTCAACCAA  
 501 AGGTAATGCG ATGGCGGCG TAGGTGCGCG CAAATGGCTG GAAAGCGGAG  
 551 CATCTGTCGG TGTGCTTAC GGGCACAGCA GGCGCAGCGT GGCGCAAAAT  
 601 TACCGCGTGG GCGGCGGCGG GCAGCACATC GGAAATTTTG GCGCGGAATA  
 50 651 TTTGGAACGG CGCAAGCAGC GATATTTGT ACAAGAGGGT GCTTTGAAAT  
 701 TCAATTCCGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG  
 751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAACTAC AaAAATACAT  
 801 CGAAGAGCAT GACAAAAGCT GGCGGGAAAA CCTg.CaCCG CAATACGACA  
 851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG  
 55 901 TTTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTTCG  
 951 CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC  
 1001 AGTTCAATTA CGGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC

5  
10  
15  
20  
25  
30

```

1051 GCAGCCTACA ATTCGGGCAG GCAGAAATAT CCGAAAGGGT CGAAGTTTAC
1101 AGGCTGGGGG CTTTAAAGG ATTTGAAAC CTACAACAAC GCGAAAATCC
1151 TCGACCTCAA CAACACCGCC ACCTTCCGGC TGCCCCGCGA AACCGAGTTG
1201 CAAACCACTT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAACCG
1251 CTTTCCTGAA GAATTGGGGC TGTTTTTCGA CGGTCCTGAT CAGGACAACG
1301 GGCTTTATTC CTATTGGGGG CGGTTTAAGG GCGATAAAGG GCTGCTGCCC
1351 CAAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATATT TCAACACGTT
1401 CTACTTCGAT GCCCGCTCA AAAAAGACAT TTACCGCTTA AACTACAGCA
1451 CCAATACCGT CGGCTACCGT TTCGGCGGCG AATATACGGG CTATTACGGC
1501 TCGGATGACG AATTTAAGCG GGCATTGCGA GAAAACTCGC CGACATACAA
1551 GAAACATTGC AACCGGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAAT
1601 ACGGCAAAAA GCGCGCCAAC AACCATTCCG TCAGCATTAG TGCGGACTTC
1651 GCGGATTATT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATGCC
1701 CAACATCCAA GAAATGTATT TTTCCCAAAT CGGCGACTCC GCGGTTTACA
1751 CCGCCTTAAA ACCAGAGCGC GCAAAACACTT GGCAATTTGG CTTCAATACC
1801 TATAAAAAAG GATTGTAAAA ACAAGATGAT ACATTAGGAT TAAAACTGGT
1851 CCGGTACCGC AGCCGCATCG ACAACTACAT CCACAACGTT TACGGGAAAT
1901 GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT
1951 GCCTACACCA TCCAACATCG CAATTTCAAA GACAAAGTGC ACAAACACGG
2001 TTTTGTAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT
2051 CTTACGCTTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC
2101 GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG
2151 GTTGAGCAGG GTTCCGCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG
2201 GTACGCGCTG GTTGGGCAAC AAACCTGACTT TGGGCGGCGC GATGCGCTAT
2251 TTCGGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC
2301 CAACGGGGGA AATACCAGCA ATTTCCGGCA ACTGGGCAAG CGTCCATCA
2351 AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTTGA TTTTACGCC
2401 GCTTACGAGC CGAAGAAAAA CCTTATTTTC CGCGCCGAAG TCAAAAATCT
2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA
2501 CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA
2551 ACGTGTAAAT CTGATAAAAC GTTGTGCAAC GGCAAATACG GCGGCACAAG
2601 CAAAAGCGTA TTGACCAATT TTGCACGCGG ACGCACCTTT TTGATGACGA
2651 TGAGCTACAA GTTTTAA

```

35 This corresponds to the amino acid sequence (SEQ ID NO: 878; ORF133-1):

40  
45  
50

```

1 EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI
51 PGAFTQQDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFYS TSTDAGRAGG
101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVVQGN
151 NTYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN
201 YRVGGGGQHI GNFGAEYLER RKQRYFVQEG ALKFNSDSGK WERDLQRQW
251 KYKPYKNYNN QELQKYEIEH DKSRENLPX QYDITPIDPS SLKQQSAGNL
301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNYGLS LNPYTNLNL
351 AAYNSGRQKY PKGSKFTGWG LLKDFETYNN AKILDNLNTA TFRLPRETEL
401 QTTLGFNYFH NEYGKNRFPE ELGLFFDGPD QDNGLYSYLG RFGDKGLLP
451 QKSTIVQPAG SQYFNTFYFD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG
501 SDDEFKRAFG ENSPTYKKHC NRSCGIYEPV LKKYGKKRAN NHSVSISADF
551 GDYFMPFASY SRTHRPNIQ EMYFSQIGDS GVHTALKPER ANTWQFGFNT
601 YKKGLLKQDD TLGLKLVGYS SRIDNYIHNV YGKWWDLNGD IPSWVSSTGL
651 AYTIOHRNFK DKVHKHGFEL ELNYDYGRFF TNLSYAYQKS TQPTNFSAS
701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGTRWLGK KLTLLGAMRY
751 FGKSIRATAE ERYIDGTNGG NTSNFRQLGK RSIKQETELA RQPLIFDFYA
801 AYEPKKNLIF RAEVKNLFDR RYIDPLDAGN DAATQRYYS FDPKDKDEDV
851 TCNADKTLN GKYGGTSKSV LTNFARGRTF LMTMSYKF*

```

55 Computer analysis of this amino acid sequence gave the following results:

Homology with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801) (SEQ ID.NO: 1167)

ORF133 (SEQ ID NO: 876) and HI121 (SEQ ID NO: 1167) show 57% aa identity in 363aa overlap:

5 Orf133: 31 IYEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90  
 I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA  
 HI121: 563 INEPILHKSGHKAFNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFQSQVSNAGVNTA 622

10 Orf133: 91 LKPERANTWQFGFXYTKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150  
 LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHNVYG WW +P+W  
 HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIHNVYGVWW--RDGMPTWA 680

Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXXYDYGFRFTNLSYAYQKSTQPTNFSDESPPN 210  
 S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN  
 HI121: 681 ESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAYQRTNQPTNYADASPRPNN 740

15 Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSIRATAEERYID 270  
 AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A RY+GKS RAT EE YI+  
 HI121: 741 ASQEDILKQGYGLSRVSMPLPKDYGRLELGTWFDQKLTGLAARYYGKSKRATIEEEYIN 800

Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDP 330  
 G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP  
 HI121: 801 GSR-FKKNLRRRENYAVKKTEDIKKQPIILDLHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

20 Orf133: 331 LDAGNDAAXERYSSFPDKDKDXDVTCTNADKTLNCGKYGGTSKSVLTNFARGRTFLMTMS 390  
 LDAGNDAA +RYSS + + C D + C GG+ K+VL NFARGRT++++++  
 HI121: 860 LDAGNDAASQRYSS- - - - -NNSIECAQDSSAC- - - - -GGSDKTVLYNFARGRTYILSLN 910

Orf133: 391 YKF 393  
 YKF  
 25 HI121: 911 YKF 913

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 (SEQ ID NO: 876) shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) (SEQ ID NO: 880) from strain A of *N. meningitidis*:

[illegible]



-603-

	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYSRIDNYIHNVYGKWWDLNGDIPSWVS
	orf133a	KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYSRIDXYIHNVYGKWWDLNGNIPSWVS
		570 580 590 600 610 620
5	orf133.pep	STGLAYTIQHRFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDASESPNNA
	orf133a	STGLAYTIQHRNFKDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSDASESPNNA
		630 640 650 660 670 680
10	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLVGTRWLGKNTLGGAMRYFGKSIRATAEERYIDG
	orf133a	SKEDQLKQGYGLSRVSALPRDYGRLVGTRWLGKNTLGGAMRYFGKSIRATAEERYIDX
		690 700 710 720 730 740
15	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL
	orf133a	TNGXXTSNFRQLGKRSIXQTETLARQPLIFDXAAYEPKKXLI FRAEVKNLFDRRYIDPL
		750 760 770 780 790 800
20	orf133.pep	DAGNDAAXERYSSFDPKDKDXDVT CNADKTL CNKYGGSVLTNFARGRTFLMTMSY
	orf133a	DAGNDAATQRYSSFDPKDKDEEVT CNDDNTL CNKYGGSVLTNFARGXTFLITMSY
		810 820 830 840 850 860
25	orf133.pep	KFX
	orf133a	KFX
		870

A partial ORF133a nucleotide sequence (SEQ ID NO: 879) is:

	1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
	51	TATATTCAAA	TCCANCGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
35	101	GTGCGTTTAC	ACANCAANAT	AAAAGCTCGG	GCNTTGTGTC	TTTGAATATT
	151	CGCNGCGACA	GCGGTTTCGG	GCGGGTCAAT	ACNATGGTNG	ACGCATCAC
40	201	NCANACCTTT	TATTCGACTT	CTACCGATGC	GGGCAGGGCA	GGCGGTTTAT
	251	CTCAATTCGG	TGCATCTGTC	GACAGCAATT	TTATNGCCGG	ACTGGATGTC
45	301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAGGC	ATCAACAGCC	TTGCCGGTTC
	351	GGCGAATCTG	CGGACTTTAN	GCGTGGATGA	TGTCGTTTCA	GGCAATANTA
50	401	CNTACGGCCT	GCTGCTAAAA	GGTCTGACCG	GCACCAATTC	AACCAAAGGT
	451	AATGCGATGG	CGGCGATAGG	TGCGCGCAAA	TGGCTGGAAA	GCGGAGCATC
	501	TGTCGGTGTG	CTTTACGGGC	ACAGCAGGCG	CAGCGTGGCG	CAAAATTACC
	551	GCGTGGGCGG	CGGCGGGCAG	CACATCGGAA	ATTTTGGCGC	GGAATATCTG
	601	GAACGACGCA	AGCAACGATA	TTTTGAGCAA	GAAGGCGGGT	TGAAATTCAA
	651	TTCCAACAGC	GGAAAATGGG	AGCGGGATTT	CCAAAAGTCG	TACTGGAAAA
	701	CCAAGTGTA	TCAAAAATAC	GATGCCCCC	AAGAACTGCA	AAAATACATC
	751	GAAGGTCATG	ATAAAAGCTG	GCGGGAAAAAC	CTGGCGCCGC	AATACGACAT
	801	CACCCCATC	GATCCGTCCA	GCCTGAAGCN	GCAGTCGGCA	GGCAACCTGT
	851	TTAAATTGGA	ATACGACGGC	GTATTCAATA	AATACACGGC	GCAATTTTCG
	901	GATTTAAACA	CCAAAATCGG	CAGCCGCAAA	ATCATCAACC	GCAATTATCA
	951	ATTCAATTAC	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AATCTGACCG
	1001	CAGCCTACAA	TTCGGGCAGG	CAGAAATATC	CGAAAGGGTC	GAAGTTTACA
	1051	GGCTGGGGGC	TTTTNAAAGA	TTTTGAAACC	TACAACAACG	CAAAAATCCT

5  
1101 CGACCTCANC AACACCTCCA CCTTCCGGCT GCCCCGTGAA ACCGAGTTGC  
1151 AAACCACTTT GGGCTTCAAT TATTTCCACA ACGAATACGG CAAAAACCGC  
1201 TTTCTGAAG AATTGGGGCT GTTTTTCGAC GGTCCGGATC ANGACAACGG  
1251 GCTTTATTCC TATTTGGGGC GGTTTAAGGG CGATAAAGGG CTGCTGCCCC  
1301 AAAAAACAAC CATTGTCCAA CCGGCCGGCA GCCAATATTT CAACACGTTT  
1351 TACTTCGATG CCGCGCTCAA AAAAGACATT TACCGCTTAA ACTACAGCAC  
1401 CAATACCGTC GGCTACCGTT TCGGCGGCNA ATATACGGGC TATTACNGCT  
1451 CGGATGACGA ATTTAAGCGG GCATTCGGAG AAAACTCGCC GACATACANG  
1501 AAACATTGCA ACCAGAGCTG CGGAATTTAT GAACCCGTAT TGAAAAAATA  
10  
1551 CGGCAAAAAG CGGCCAACA ACCATTCCGT CAGCATTAGT GCGGACTTCG  
1601 GCGATTATTT CATGCCGTTT GCCAGCTATT CGCGCACACA CCGTATGCCC  
1651 AACATCCAAG AAATGTATTT TTCCCAAATC GGCGACTCCG GCGTTCACAC  
1701 CGCCTTAAAA CCAGAGCGCG CAAACACTTG GCAATTTGGC TTCAATACCT  
1751 ATAAAAAAGG ATTGTTAAAA CAAGATGATA TATTAGGATT AAAACTGGTC  
15  
1801 GGCTACCGCA GCCGCATCGA CNACTACATC CACAACGTTT ACGGGAAATG  
1851 GTGGGATTTG AACGGGAATA TTCCGAGCTG GGTCAGCAGC ACCGGGCTTG  
1901 CCTACACCAT CCAACACCGC AATTTCAAAG ACAAAGTGCA CAAACACGGT  
1951 TTTGAGTTGG AGCTGAATTA CGATTATNGG CGTTTTTTCA CCAACCTTTC  
2001 TTACGCCTAT CAAAAAGCA CGCAACCGAC CAACTTCAGC GATGCGAGCG  
20  
2051 AATCGCCCAA CAATGCGTCC AAAGAAGACC AACTCAAACA AGGTTATGGG  
2101 TTGAGCAGGG TTCCGCCCTT GCCGCGAGAT TACGGACGTT TGGAAGTCGG  
2151 TACGCGCTGG TTGGGCAACA AACTGACTTT GGGCGGCGCG ATGCGCTATT  
2201 TCGGCAAGAG CATCCGCGCG ACGGCTGAAG AACGCTATAT CGACGNCACC  
2251 AATGGGGNAN NTACCAGCAA TTCCGGCAA CTGGGCAAGC GTTCCATCAN  
25  
2301 ACAAACCGAA ACCCTTGCCC GCCAGCCTTT GATTTTTGAT TTNTACGCCG  
2351 CTTACGAGCC GAAGAAAAAN CTTATTTTCC GCGCCGAAGT CAAAAATCTG  
2401 TTCGACAGGC GTTATATCGA TCCGCTCGAT GCGGGCAATG ATGCGGCAAC  
2451 GCAGCGTTAT TACAGTTCGT TCGACCCGAA AGACAAGGAC GAAGAAGTAA  
2501 CGTGTAATGA TGATAACAG TTATGCAACG GCAAATACGG CGGCACAAGC  
30  
2551 AAAAGCGTAT TGACCAATTT TGCACGCGGA CNACCTTTT TGATAACGAT  
2601 GAGCTACAAG TTTTAA

This encodes a protein having (partial) amino acid sequence (SEQ ID NO: 880):

35  
1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI  
51 RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV  
101 VKGSFSGSAG INSLAGSANL RTLXVDDVQ GNXTYGLLLK GLTGTNSTKG  
151 NAMAAGARK WLESGASGV LYGHSRRSVA QNYRVGGGQ HIGNFGAEYL  
201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTKWYQKY DAPQELQKYI  
40  
251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR  
301 DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT  
351 GWGLXKDFET YNNAKILDLY NTSTFRLPRE TELQTTLGFN YFHNEYGKNR  
401 FPEELGLFFD GPDXDNGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF  
451 YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX  
501 KHCNQSCGIY EPVLKKGK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP  
45  
551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLK  
601 GYRSRIDXYI HNVYKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG  
651 FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG  
701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT  
751 NGXXTGNFRQ LGKRSIXQTE TLRQPLIFD XYAAYEPKXK LIFRAEVKNL  
50  
801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS  
851 KSVLTNFARG XTFLITMSYK F\*

ORF133a (SEQ ID NO: 880) and ORF133-1 (SEQ ID NO: 878) show 94.3% identity in 871 aa overlap:

55  
orf133a.pep  
10 20 30 40  
KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS

	orf133-1	EAQIQVLEDVHVKA	KRV	PKDKK	VFTD	ARAV	STRQ	DIFK	SS	ENLD	NIVRS	IPGA	FTQ	QDKS
		10	20	30	40	50	60							
5	orf133a.pep	SGXVSLNIRXDS	SGFGRVNTM	VDGITXTFY	STSTDA	GRAGGSSQ	FGASVDS	NFXAGL	DVVK					
	orf133-1	SGIVSLNIRGDS	SGFGRVNTM	VDGITQTFY	STSTDA	GRAGGSSQ	FGASVDS	NFIAGL	DVVK					
10	orf133a.pep	GSFSGSAGINSL	AGSANLRTL	XVDDVQGN	XTYGLLL	KGLTGTN	STKGNAMA	AAIGARK	WL					
	orf133-1	GSFSGSAGINSL	AGSANLRTL	GVDDVQGN	NNTYGLLL	KGLTGTN	STKGNAMA	AAIGARK	WL					
15	orf133a.pep	ESGASVGVLYG	HSSR	SV	AQNYRV	GGGGQH	IGNFGAEY	LERRKQ	RYFEQ	EGGLKF	NSNSGK			
	orf133-1	ESGASVGVLYG	HSSR	SV	AQNYRV	GGGGQH	IGNFGAEY	LERRKQ	RYFVQ	EGALKF	NSDSGK			
20	orf133a.pep	WERDFQKSYW	KTKWYQ	KYDAPQ	ELQKYI	EGHDKS	WRENLP	QYDITP	IDPSSL	KXQSAG	N			
	orf133-1	WERDLQRQQW	KYKPYK	NYNN-QE	LQKYIE	EHDKS	WRENLP	QYDITP	IDPSSL	KXQSAG	N			
25	orf133a.pep	LFKLEYDGVF	NKYTAQ	FRDLN	TKIGSR	KIINRNY	QFNYGL	SLNPYT	NLNLTA	AYNSGR	QK			
	orf133-1	LFKLEYDGVF	NKYTAQ	FRDLN	TKIGSR	KIINRNY	QFNYGL	SLNPYT	NLNLTA	AYNSGR	QK			
30	orf133a.pep	YPKGSKFTG	WGLXK	DFETYN	NAKILD	LXNTST	FRLPRE	TELQTT	LGFN	YFHNE	YGKNRFP			
	orf133-1	YPKGSKFTG	WGLXK	DFETYN	NAKILD	LXNTST	FRLPRE	TELQTT	LGFN	YFHNE	YGKNRFP			
35	orf133a.pep	EELGLFFD	GPDND	GLYSYL	GRFKG	DKGLLP	QKSTIV	QPAGS	QYFNT	FYFDA	AALKK	DIYR		
	orf133-1	EELGLFFD	GPDND	GLYSYL	GRFKG	DKGLLP	QKSTIV	QPAGS	QYFNT	FYFDA	AALKK	DIYR		
40	orf133a.pep	LNYSNTVGY	RFGGXY	TGYXSD	DEFKRA	FGENS	PTYXKH	CNQSC	GIYEP	VLKKG	YKKRA			
	orf133-1	LNYSNTVGY	RFGGEY	TGYGSD	DEFKRA	FGENS	PTYXKH	CNRSC	GIYEP	VLKKG	YKKRA			
45	orf133a.pep	NNHSVSI	SADFGDY	FMPFAS	YSRTH	RMPN	IQEMYF	SQIGD	SGVHT	ALKPER	ANTWQ	FGFN		
	orf133-1	NNHSVSI	SADFGDY	FMPFAS	YSRTH	RMPN	IQEMYF	SQIGD	SGVHT	ALKPER	ANTWQ	FGFN		

		590	600	610	620	630	640
	orf133a.pep	TYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVSSTGLAYTIQHRNF					
5	orf133-1	TYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVSSTGLAYTIQHRNF	600	610	620	630	640
		650	660	670	680	690	700
	orf133a.pep	KDKVHKHGFEELELNLDYXRFFTNLSYAYQKSTQPTNFSASESPNNASKEDQLKQGYGLS					
10	orf133-1	KDKVHKHGFEELELNLDYGRFFTNLSYAYQKSTQPTNFSASESPNNASKEDQLKQGYGLS	660	670	680	690	700
		710	720	730	740	750	760
	orf133a.pep	RVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG					
15	orf133-1	RVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLG	720	730	740	750	760
		770	780	790	800	810	820
	orf133a.pep	KRSIXQTETLARQPLIFDXAAYEPKXLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS					
20	orf133-1	KRSIXQTETLARQPLIFDFYAAYEPKKNLIFRAEVKNLFDRRYIDPLDAGNDAATQRYYS	780	790	800	810	820
		830	840	850	860	870	
	orf133a.pep	SFDPKDKDEEVTNCDDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
25	orf133-1	SFDPKDKDEDVTNCADKTLNCGKYGGTSKSVLTNFARGRTFLMTMSYKFX	840	850	860	870	880

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF133 (SEQ ID NO: 876) shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) (SEQ ID NO: 882) from *N. gonorrhoeae*:

30	orf133.pep	PGYYGSDDEFKRAFGENSPXKKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
	orf133.pep	YEPVLKKYGKKRANNHSVSIADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGKKRANNHSVSIADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
35	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
	orf133.pep	STGLAYTIQHRXFXDKVHGXQXXXXXXXXYDGRFFTNLSYAYQKSTQPTNFSASESPNNA	211
40	orf133ng	STGLAYTIRHRNFKDKVHKHGFEELELNLDYGRFFTNLSYAYQKSTQPTNFSASESPNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLVGTRWLGKLTGGAMRYFGKSIRATAEERYIDG	800

orf133.pep TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL 331  
 orf133ng TNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLIFRAEVKNLFDRRYIDPL 860

5 orf133.pep DAGNDAAXERYSSFDPKDKDXDVTNCNADKTLKNGKYGGSVLTNFARGRTFLMTMSY 391  
 orf133ng DAGNDAATQRYSSFDPKDKDEDVTCNADKTLKNGKYGGSVLTNFARGRTFLMTMSY 920

orf133.pep KF 393  
 orf133ng KF 922

The complete length ORF133ng nucleotide sequence (SEQ ID NO: 881) is predicted to encode a protein having amino acid sequence (SEQ ID NO: 882):

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAIQIVL EDVHVKAQRV  
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN  
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD  
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDV VQNNYTGILL KGLTGTNSTK  
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY  
 251 LERRKQYFV QEGGLKFENAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY  
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLLNLEYD GVFNKYTAQF  
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKE  
 401 TGWGLLKDFE TYNNAKILDL NNTATFRLPR ETELQTTLGF NYFHNEYGKN  
 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT  
 501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYYGSENEFK RAFGENSPAY  
 551 KEHCDPSCGL YEPVLKKYK KKRANNHVSIS SADFGDYFMP FAGYSRTHRM  
 601 PNIQEMYFSQ IGDGSHVTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL  
 651 VGYRSRIDNY IHNVYKWDW LNGDIPSWVG STGLAYTIRH RNFKDKVHKH  
 701 GFELELNLDY GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY  
 751 GLSRVSLPR DYGRLEVGTG WLGKLTLLGG AMRYFGKSIR ATAEERYIDG  
 801 TNGGNTSNVR QLGRSIKQT ETLARQPLIF DFYAAAYEPK NLI FRAEVKN  
 851 LFDRLRYIDPL DAGNDAATQR YSSFDPKDK DEDVTCNADK TLCNGKYGGS  
 901 SKSVLTNFAR GRTFLMTMSY KF\*

A variant was also identified, being encoded by the gonococcal DNA sequence (SEQ ID NO: 883):

1 ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGT'TTTTATC TTATGGGTGT  
 51 TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG  
 101 AGGCGCAGAT ACAGGTTTGT GAAGATGTGC ACGTCAAGGC GAAGCGCGTA  
 151 CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca  
 201 gGATGTGTTC AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC  
 251 CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATTGT GTCTTTGAAT  
 301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT  
 351 CACGCAGACC TTTTATTCGA CTTCTACCGA TGCGGGCAGG GCAGGCGGTT  
 401 CATCTCAATT CGGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT  
 451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG  
 501 TTCGGCGAAT CTGCGGACTT TAGGCGTGGA TGACGTCGTT CAGGGCAATA  
 551 ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA  
 601 GGTAATGCGA TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC  
 651 GTCTGTCTGGT GTGCTTTACG GGCACAGCAG GCGGCGCGTG GCGCAAAATT  
 701 ACCGCGTGGG CGGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAATAT  
 751 CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT  
 801 CAATGCCGGC AGCGGAAAAT GGGAACGGGA TTTGCAAAGG CAATACTGGA  
 851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCCAAGAACT GCAAAAAATAC  
 901 ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA  
 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC

	1001	TGTTTAAATT	GGAATACGAC	GGCGTATTCA	ATAAATACAC	GGCGCAATTT
	1051	CGCGATTTAA	ACACCAGAAT	CGGCAGCCGC	AAAATCATCA	ACCGCAATTA
	1101	TCAATTCAAT	TACGGTTTGT	CTTTGAACCC	GTATACCAAC	CTCAATCTGA
5	1151	CCGCAGCCTA	CAATTCGGGC	AGGCAGAAAT	ATCCGAAAGG	GGCGAAGTTT
	1201	ACAGGCTGGG	GGCTTTTAAA	AGATTTTGAA	ACCTACAACA	ACGCGAAAAT
	1251	CCTCGACCTC	AACAACACCG	CCACCTTCCG	GCTGCCCCGC	GAAACCGAGT
	1301	TGCAAACCAC	TTTGGGCTTC	AATTATTTCC	ACAACGAATA	CGGCAAAAAC
	1351	CGCTTTCCTG	AAGAATTGGG	GCTGTTTTTC	GACGGTCCTG	ATCAGGACAA
	1401	CGGGCTTTAT	TCCTATTTGG	GGCGGTTTAA	GGGCGATAAA	GGGCTGTTGC
10	1451	CTCAAAAATC	AACCAATGTC	CAACCGCCCG	GCAGCCAATA	TTTCAACACG
	1501	TTCTACTTCG	ATGCCGCGCT	CAAAAAAGAC	ATTTACCGCT	TAAACTACAG
	1551	CACCAATGCA	ATCAACTACC	GTTTCGGCGG	CGAATATACG	GGCTATTACG
	1601	GCTCGGAAAA	CGAATTTAAG	CGGGCATTCG	GAGAAAACCTC	GCCCGCATAC
	1651	AAGGAACATT	GCGACCCGAG	CTGCGGGCTT	TATGAACCCG	TATTGAAAAA
15	1701	ATACGGCAAA	AAGCGCGCCA	ACAACCATT	GGTCAGCATT	AGTGGCGACT
	1751	TCGGCGATTA	TTTCATGCCG	TTCGCCGCGT	ATTCGCGCAC	ACACCGTATG
	1801	CCCAACATCC	AAGAAATGTA	TTTTTCCCAA	ATCGGCGACT	CCGGCGTTCA
	1851	CACCGCCTTA	AAACCAGAGC	GCGCAAACAC	TTGGCAATTT	GGCTTCAATA
	1901	CCTATAAAAA	AGGATTGTTA	AAACAAGATG	ATATATTAGG	ATTGAACTG
20	1951	GTCGGCTACC	GCAGCCGCAT	TGACAACCTAC	ATCCACAACG	TTTACGGGAA
	2001	ATGGTGGGAT	TTGAACGGGG	ATATTCGGAG	CTGGGTCGGC	AGCACCGGGC
	2051	TTGCCTACAC	CATCCGACAC	CGCAATTTC	AAGACAAAGT	GCACAAACAC
	2101	GGTTTTGAGC	TGGAGCTGAA	TTACGATTAT	GGGCGTTTTT	TCACCAACCT
	2151	TTCTTACGCC	TATCAAAAAA	GCACGCAACC	GACCAATTTT	AGCGATGCGA
25	2201	GCGAATCGCC	CAACAATGCC	tccaaAGAAG	ACCAACTCAA	ACAAGGTTAT
	2251	GGGCTGAGCA	GGGTTTCCGC	CCTGCCGCGA	GATTACGGAC	GTTTGGAAGT
	2301	CGGTACGCGC	TGGTTGGGCA	ACAACTGAC	TTTGGGCGGC	GCGATgcGCT
	2351	ATTTTCGGCAA	GAGCATCCGC	GCGACGGCTG	AAGAACGCTA	TATCGACGGC
	2401	ACCAACGGGG	GAAATACCAG	CAATGTCCGG	CAACTGGGCA	AGCGTTCAT
30	2451	CAAAACAAAC	GAAACCTTG	CCCGACAGCC	TTTGATTTTT	GATTTTTACG
	2501	CCGCTTACGA	GCCGAAGAAA	AACCTTATTT	TCCGCGCCGA	AGTCAAAAAC
	2551	CTGTTTCGACA	GGCGTTATAT	CGATCCGCTC	GATGCGGGCA	ATGATGCGGC
	2601	AACGCAGCGT	TATTACAGCT	CGTTCGACCC	GAAAGACAAG	GACGAAGACG
	2651	TAACGTGTAA	TGCTGATAAA	ACGTTGTGCA	ACGGCAAATA	CGGCGGCACA
35	2701	AGCAAAAGCG	TATTGACCAA	TTTCGCACGC	GGACGCACCT	TCTTGATGAC
	2751	GATGAGCTAC	AAGTTTTTAA			

This corresponds to the amino acid sequence (SEQ ID NO: 884; ORF133ng-1):

	1	MRSSFRLKPI	CFYLMGVMLY	HHSYAEDAGR	AGSEAQIQVL	EDVHVKAKR
40	51	PKDKKVFTDA	RAVSTRQDVF	KSGENLDNIV	RSIPGAFTQQ	DKSSGIVSLN
	101	IRGDSGFGRV	NTMVDGITQT	FYSTSTDAGR	AGGSSQFGAS	VDSNFIAGLD
	151	VVKGSFSGSA	GINSLAGSAN	LRTLGVDDVV	QGNNTYGLLL	KGLTGTNSTK
	201	GNAMAAIGAR	KWLESASVG	VLYGHSRRGV	AQNYRVGGGG	QHIGNFGEEY
	251	LERRKQYFV	QEGGLKFNAG	SGKWERDLQR	QYWTKWYKK	YEDPQELQKY
45	301	IEEHDKSWRE	NLAPQYDITP	IDPSGLKQQS	AGNLFKLEYD	GVFNKYTAQF
	351	RDLNTRIGSR	KIINRNYQFN	YGLSLNPYTN	LNLTAAYNSG	RQKYPKGAKF
	401	TGWLLKDFE	TYNNAKILDL	NNTATFRLPR	ETELQTTLGF	NYFHNEYGKN
	451	RFPEELGLFF	DGPDQDNGLY	SYLGRFKGDK	GLLPQKSTIV	QPAGSQYFNT
	501	FYFDAALKKD	IYRLNSTNA	INYRFGGEYT	GYGSENEFK	RAFGENSPAY
50	551	KEHCDPSCGL	YEPVLKKYK	KRANNHSVSI	SADFGDYFMP	FAGYSRTHRM
	601	PNIQEMYFSQ	IGDSGVHTAL	KPERANTWQF	GFNTYKKGLL	KQDDILGLKL
	651	VGYSRIDNY	IHNVYGKWD	LNGDIPSWVG	STGLAYTIRH	RNFKDKVHKH
	701	GFELELNIDY	GRFFTNLSYA	YQKSTQPTNF	SDASESPNNA	SKEDQLKQGY
	751	GLSRVSALPR	DYGRLEVGT	WLGKLTLLGG	AMRYFGKSIR	ATAEERYIDG
55	801	TNGGNTSNVR	QLGKRSIKQT	ETLARQPLIF	DFYAAIEPKK	NLIFRAEVKN
	851	LFDRRYIDPL	DAGNDAATQR	YYSSFDPKDK	DEDVTCNADK	TLCNGKYGGT
	901	SKSVLTNFR	GRTFLMTMSY	KF*		

		10	20	30	40	50	60
5	orf133ng-1.pep	SFRLK	PICFYLMGVMLYHHSYAEDAGRAGSE	AEQIQVLEDVHV	KAKRVPKDKK	VFTDARAV	
	orf133-1			EAQIQVLEDVHV	KAKRVPKDKK	VFTDARAV	
				10	20	30	
		70	80	90	100	110	120
10	orf133ng-1.pep	STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS					
	orf133-1	STRQDIFKSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS					
		40	50	60	70	80	90
		130	140	150	160	170	180
15	orf133ng-1.pep	TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN					
	orf133-1	TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN					
		100	110	120	130	140	150
		190	200	210	220	230	240
20	orf133ng-1.pep	NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGHSSRRGVAQNYRVGGGGQHI					
	orf133-1	NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLVYGHSSRRSVAQNYRVGGGGQHI					
		160	170	180	190	200	210
		250	260	270	280	290	300
25	orf133ng-1.pep	GNFGEEYLERRKQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWKYKYPQELQKYIEE					
	orf133-1	GNFGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQRQQWKYKPKYNYN-QELQKYIEE					
		220	230	240	250	260	
		310	320	330	340	350	360
30	orf133ng-1.pep	HDKSWRENAPQYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII					
	orf133-1	HDKSWRENLPQYDITPIDPSSLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKII					
		270	280	290	300	310	320
		370	380	390	400	410	420
35	orf133ng-1.pep	NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDLNNT					
	orf133-1	NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDLNNT					
		330	340	350	360	370	380
		430	440	450	460	470	480
40	orf133ng-1.pep	ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPDQDNGLYSYLGRFGKDGKLL					
	orf133-1	ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPDQDNGLYSYLGRFGKDGKLL					
		390	400	410	420	430	440
		490	500	510	520	530	540
45	orf133ng-1.pep	PQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAF					
	orf133-1	PQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSTNTVGYRFGGEYTGYYGSDDEFKRAF					
		450	460	470	480	490	500

-610-

5  
10  
15  
20  
25  
30  
35

```

      550      560      570      580      590      600
orf133ng-1.pep GENSPAYKEHCDPSCGLYEPVLKKYGGKRRNNHSVSISADFGDYFMPFAGYSRTHRMPNI
|||::||::||: ||::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       GENSPTYKKHCNRS CGIYEPVLKKYGGKRRNNHSVSISADFGDYFMPFASYSRTHRMPNI
      510      520      530      540      550      560

      610      620      630      640      650      660
orf133ng-1.pep QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHN
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN
      570      580      590      600      610      620

      670      680      690      700      710      720
orf133ng-1.pep VYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVKHKGFELELNVDYGRFFTNLSYAYQK
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       VYGKWWDLNGDIPSWVSSTGLAYTIQHRNFKDKVKHKGFELELNVDYGRFFTNLSYAYQK
      630      640      650      660      670      680

      730      740      750      760      770      780
orf133ng-1.pep STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLGNKLT LGGAMR
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLGNKLT LGGAMR
      690      700      710      720      730      740

      790      800      810      820      830      840
orf133ng-1.pep YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       YFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
      750      760      770      780      790      800

      850      860      870      880      890      900
orf133ng-1.pep FRAEVKNLFD RRYIDPLDAGNDAATQRYYS SFDPKDKDEDVTCNADKTL CNGKYGGT SKS
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       FRAEVKNLFD RRYIDPLDAGNDAATQRYYS SFDPKDKDEDVTCNADKTL CNGKYGGT SKS
      810      820      830      840      850      860

      910      920
orf133ng-1.pep VLTNFARGRTFLMTMSYKFX
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1       VLTNFARGRTFLMTMSYKFX
      870      880

```

In addition, ORF133ng-1 (SEQ ID NO: 884) is homologous to a TonB-dependent receptor (SEQ ID NO: 1167) in *H.influenzae*:

40  
45

```

sp|P45114|YC17_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR
)gi|1075372|pir||G64110 transferrin binding protein 1 precursor (tbpl) homolog -
Haemophilus influenzae (strain Rd KW20) )gi|1574147 (U32801) transferrin binding
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913
Score = 930 bits (2377), Expect = 0.0
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVKA KRVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKGSGVV 88

Query: 98 SLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVKG SFS 157

```



S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS  
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSQFGAAIDPNFIAGVDVNKSNS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217  
G++GIN+LAGSAN RTLGV+DV+ M RKWL++G  
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLNDGG 208

Query: 218 SVGVLGHSSRRGVAQNYRVGGGQHIGNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERD 277  
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D  
Sbjct: 209 YGVVYGYSQREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265

Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303  
L +++W +Y KK +D ++LQK IEE  
Sbjct: 266 LSKKHWSCNKPQYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIEE 325

Query: 304 HDKSWRENLAQYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363  
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI  
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQLGAQLRTLNDKIGSRKIE 384

Query: 364 NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT 423  
NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+  
Sbjct: 385 NRNYQVNYNFNNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

Query: 424 ATFRLPRETELQTTLGFNYFHNEYGKNRFPPEELGLFFDGPQDQNGLYSY--LGRFKGDKG 481  
TF LP+E +L+TTLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K  
Sbjct: 445 HTFLLPKEIDLKTTLGFNYFTNEYSKNRFPPEELSLFYNDASHDQGLYSHSKRGRYSGTKS 504

Query: 482 LLPQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541  
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY  
Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDLTALSGIYHLNYSVNFTHYAFNGEYVGY----- 555

Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKGRANNHVSISADFGDYFMPFAGYSRTHRMP 601  
EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP  
Sbjct: 556 ---ENTAGQQ-----INEPILHKSGHKKAFNHSATLSAELSDYFMPFFTYSRTHRMP 604

Query: 602 NIQUEMYSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661  
NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI  
Sbjct: 605 NIQUEMFSQVSNAGVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIYI 664

Query: 662 HNVYGKWDLNGDIRSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 721  
HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY  
Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAY 722

Query: 722 QKSTQPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGA 781  
Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTG A  
Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGTRWFDQKLTGLA 782

Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKN 841  
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+  
Sbjct: 783 ARYYGKSKRATIEEEYINGSR-FKKNLRLRENYAVKKTEDIKKQPIILDHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYYSSFPKDKDEDVTCNADKTLNKGKYGTS 901  
LI +AEV+NL D+RY+DPLDAGNDAA+QRYYSS + + C D + C GG+  
Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYYSS-----NNSIECAQDSSAC---GGSD 892

Query: 902 KSVLTNFARGRTFLMTMSYKF 922  
K+VL NFARGRT+++++YKF  
Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* (SEQ ID NO: 885)

```

1  ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
15 401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

```

This corresponds to the amino acid sequence (SEQ ID NO: 886; ORF112):

```

1  MNLISRYIIR QMAVMVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
20 51  GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTFKLL
101 LILSQGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
151 KEKNSVINVR EMLPDH...

```

Further work revealed further partial nucleotide sequence (SEQ ID NO: 887):

```

25 1  ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 gGCTACACCG CCCTCAAAAT GCCCGCCCGC GCCTACGAAC TGATTCCCCT
30 201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCrTkAT CAATGTGCGC GAAATGTTGC CCGACCATAC
35 501 GCTTTTGGGC ATCAAAATTT GGGCGCGCAA CGATAAAAAC GAATTGGCAG
551 AGGCAGTGGG AGCCGATTCC GCCGTTTTGA ACAGCGACGG CAGTTGGCAG
601 TTGAAAAACA TCCGCCGCGC CACGCTTGGC GAAGACAAAG TCGAGGTCTC
651 TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
40 751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
801 CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC
851 TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
901 TTAAAACTCT TCGGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCGG
45 951 ACGGCTCTTT GGGTTTACCA GCCAACTCGG...

```

This corresponds to the amino acid sequence (SEQ ID NO: 888; ORF112-1):

1 MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML  
 51 GYTALKMPAR AYELIPLAVL IGGLVLSLSQL AAGSELTVIK ASGMSTKKLL  
 101 LILSQGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL  
 151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ  
 201 LKNIRRSTLG EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGELTT  
 251 YIRHLQNSQ NTRIYAIWW RKLVPAAAW VMALVAFaft PQTRRHGMMG  
 301 LKLFGGICXG LLFHLAARLF GFTSQL...

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 (SEQ ID NO: 886) shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) (SEQ ID NO: 890) from strain A of *N. meningitidis*:

15	orf112.pep	10	20	30	40	50	60
		MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
	orf112a	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR					
		10	20	30	40	50	60
20	orf112.pep	70	80	90	100	110	120
		AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLLILSQGFIFAIATVALGEW					
	orf112a	AYELMPLAVLIGGLVXSXSLAAGSELXVIKASGMSTKKLLLLILSQGFIFAIATVALGEW					
		70	80	90	100	110	120
25	orf112.pep	130	140	150	160		
		VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH					
	orf112a	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN					
		130	140	150	160	170	180
30	orf112a	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP					
		190	200	210	220	230	240

The ORF112a nucleotide sequence (SEQ ID NO: 889) is:

1 ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT  
 51 TTACGCGCTC CTTGCCTTCC TCGCTTGTA CAGCTTTTTT GAAATCCTGT  
 101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGNTG  
 151 GGNTACACCG CCCTCAAAAT GNCCGCCCGC GCCTACGAAC TGATGCCCTT  
 201 CGCCGTCCTT ATCGGCGGAC TGGTCTCTNT CAGCCAGCTT GCCGCCGCA  
 251 GCGAACTGAN CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG  
 301 TTGATTCTGT CGCAGTTCGG TTTTATTTTT GCTATTGCCA CCGTCGCGCT  
 351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG  
 401 CCGCGGCCAT CAACGCCAAA ATCAGTACCG GCAATACCGG CCTTTGGCTG  
 451 AAAGAAAAAA ACAGCATTAT CAATGTGCGC GAAATGTTGC CCGACCATAC  
 501 CCTGCTGGGC ATTAATAATCT GGGCCCGCAA CGATAAAAC GAACTGGCAG  
 551 AGGCAGTGGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG  
 601 TTGAAAAACA TCCGCCGAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC  
 651 TATTGCGGCT GAAGAAAANT GGCCGATTTC CGTCAAACGC AACCTGATGG  
 701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC



-615-

```

orf112ng      MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR  60
orf112.pep    AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW 120
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf112ng      AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQFGFIFAAVAALGEW 120
5  orf112.pep    VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH 166
               |||||:|||||:|||||:|||||:|||||:|||||:|||||
orf112ng      VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN 180

```

The complete length ORF112ng nucleotide sequence (SEQ ID NO: 891) is:

```

10      1  ATGAACCTGA  TTTCACGTTA  CATCATCCGC  CAAATGGCGG  TTATGGCGGT
        51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
       101  ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAAATGCTG
       151  GGCTACACCG  CCCTCAAAAT  GCCCGCCCGC  GCCTACGAAC  TCATGCCCTT
       201  CGCCGTCCTC  ATCGGCGGAC  TGGCCTCTCT  CAGCCAGCTT  GCCGCCGGCA
       15  251  GCGAACTGGC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
       301  TTGATTCTGT  CTCAGTTTCG  TTTTATTTT  GCTATTGCCG  CCGTCGCGCT
       351  CGGCGAATGG  GTTGCGCCCA  CGCTGAGCCA  AAAAGCCGAA  AACATCAAag
       401  cCGCCGCCAt  taacggCAAA  ATCAGCAccg  gcAATACCGG  CCTTTggcTG
       451  AAAGAAAAAa  ccAGCATTAT  CAATGTGcGc  GGAATGTTGC  CCGACCATAC
       20  501  GCTTTTGGGC  ATCAAAATTT  GGGCGCGCAA  CGATAAAAAA  GAATTGGCAG
       551  AGGCAGTGGA  AGCCGATTCC  GCCGTTTGA  ACAGCGACGG  CAGCTGGCAG
       601  TTGAAAAACA  TCCGCCGCAG  CATCATGGGT  ACAGACAAAA  TCGAAAAATC
       651  cgCCGCCGCC  GAAGAACTT  gGCCGATTGC  CGTCAGACGC  AACCTGATGG
       701  ACGTATTGCT  CGTCAAGCCC  GACCAAATGT  CCGTCGGCGA  GCTGACCACC
       25  751  TACATCCGCC  ACCTCCAAAA  CAACAGCCAA  AACACCCAAA  TCTACGCCAT
       801  CGCATGGTGG  CGTAAACTCG  TTTACCCCGT  CGCCGCATGG  GTCATGGCGC
       851  TCGTTGCCTT  CGCCTTTACG  CCGCAAACCA  CGCGCCACGG  CAATATGGGC
       901  TTAAAACTCT  TCGGCGGCAT  CTGTCTCGGA  TTGCTGTTCC  ACCTTGCCGG
       951  CAGGCTCTTC  GGGTTTACCA  GCCAACTCTA  CGGCACCCCA  CCCTTCCTCG
       30  1001  CCGGCGCACT  GCCTACCATA  GCCTTCGCCT  TGCTCGCTGT  TTGGCTGATA
       1051  CGCAAACAGG  AAAAACGTTG  A

```

This encodes a protein having amino acid sequence (SEQ ID NO: 892):

```

35      1  MNLISRYIIR  QMAVMNAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
        51  GYTALKMPAR  AYELMPLAVL  IGGLASLSQL  AAGSELAVIK  ASGMSTKKLL
       101  LILSQFGFIF  AIAAVALGEW  VAPTLSQKAE  NIKAAAINGK  ISTGNTGLWL
       151  KEKTSIINVR  GMLPDHTLLG  IKIWARNDKN  ELAEAVEADS  AVLNSDGSWQ
       201  LKNIRRSIMG  TDKIETSAAA  EETWPIAVRR  NLMDVLLVKP  DQMSVGELTT
       251  YIRHLQNSQ  NTQIYIAIWW  RKLVPVAAW  VMALVAFAPT  PQTRRHNGM
       301  LKLFGGICLG  LLFHLAGRLF  GFTSQLYGTP  PFLAGALPTI  AFALLAVWLI
       351  RKQEKR*

```

ORF112ng (SEQ ID NO: 892) and ORF112-1 (SEQ ID NO: 888) show 94.2% identity in 326 aa overlap:

```

45      orf112ng      MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
       orf112-1      MNLISRYIIRQMAVMNAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
               10      20      30      40      50      60

```

50

		70	80	90	100	110	120
	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQFGFIFAIAAVALGEW					
	orf112-1	AYELIPLAVLIGGLVSLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIAATVALGEW					
5		70	80	90	100	110	120
	orf112ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRLPDHTLLGIKIWARNDKN					
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN					
10		130	140	150	160	170	180
	orf112ng	ELAEAVEADSAVLNSDGSWQLKNIRRSIMGTDKIETSAAAEETWPIAVRRNLMVDLLVKP					
	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEENWPISVKRNLMVDLLVKP					
15		190	200	210	220	230	240
	orf112ng	DQMSVGELTTYIRHLQNSQNTQIYIAIWWRKLVYPVAAWVMALVAFATPQTTRHGNMG					
	orf112-1	DQMSVGELTTYIRHLQNSQNTRIYIAIWWRKLVYPAAWVMALVAFATPQTTRHGNMG					
20		250	260	270	280	290	300
	orf112ng	LKLFGGICLGLLFHLAGRLFGFTSQLYGTTPFLAGALPTIAFALLAVWLIRKQEKRX					
	orf112-1	LKLFGGICXGLLFHLAGRLFGFTSQL					
25		310	320	330	340	350	

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 105

- 30 Table III lists several *Neisseria* strains which were used to assess the conservation of the sequence of ORF 4 (SEQ ID NO: 216) among different strains.

TABLE III – List of *Neisseria* Strains Used for Gene Variability Study of ORF 4 (SEQ ID NO: 216)

ORF4 gene variability: List of used <i>Neisseria</i> strains		
IdentificationStrains number	Source / reference	
	<b>Group B</b>	
zv01_4	NG6/88	R. Moxon / Seiler et al., 1996
zv02_4	BZ198	R. Moxon / Seiler et al., 1996
zv03_4ass	NG3/88	R. Moxon / Seiler et al., 1996
zv04_4	297-0	R. Moxon / Seiler et al., 1996

zv05_4	1000	R. Moxon / Seiler et al., 1996
zv06_4	BZ147	R. Moxon / Seiler et al., 1996
zv07_4	BZ169	R. Moxon / Seiler et al., 1996
zv08_4	528	R. Moxon / Seiler et al., 1996
zv09_4	NGP165	R. Moxon / Seiler <i>et al.</i> , 1996
zv10_4	BZ133	R. Moxon / Seiler <i>et al.</i> , 1996
zv11_4	NGE31	R. Moxon / Seiler et al., 1996
zv12_4ass	NGF26	R. Moxon / Seiler et al., 1996
zv13_4	NGE28	R. Moxon / Seiler et al., 1996
zv15_4	SWZ107	R. Moxon / Seiler et al., 1996
zv16_4	NGH15	R. Moxon / Seiler et al., 1996
zv17_4	NGH36	R. Moxon / Seiler et al., 1996
zv18_4	BZ232	R. Moxon / Seiler et al., 1996
zv19_4	BZ83	R. Moxon / Seiler et al., 1996
zv20_4	44/76	R. Moxon / Seiler et al., 1996
zv21_4	MC58	R. Moxon
zv96_4	2996	Our collection
<b>Group A</b>		
zv22_4	205900	R. Moxon
z2491_4	Z2491	R. Moxon / Maiden et al., 1998
<b>Group C</b>		
zv24_4	90/18311	R. Moxon
zv25_4	93/4286	R. Moxon
<b>Others</b>		
zv26_4ass	A22 (group W)	R. Moxon / Maiden et al., 1998
zv27_4	E26 (group X)	R. Moxon / Maiden et al., 1998
zv28_4	860800 (group Y)	R. Moxon / Maiden et al., 1998
zv29_4	E32 (group Z)	R. Moxon / Maiden et al., 1998
<b>Gonococcus</b>		
zv32_4Ng F62	R. Moxon / Maiden <i>et al.</i> , 1998	
zv33_4	Ng SN4	R. Moxon
fa1090_4	FA1090	R. Moxon
<b><u>References:</u></b>		
Seiler A. <i>et al.</i> , Mol. Microbiol., 1996, 19(4):841-856.		
Maiden et al., Proc. Natl. Acad. Sci. USA, 1998, 95:3140-3145.		

The amino acid sequences for each listed strain are as follows:

>FA1090\_4 (SEQ ID NO: 893)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVK  
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEAF  
QVPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
KADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAK\*

>Z2491\_4 (SEQ ID NO: 894)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV01\_4 (SEQ ID NO: 895)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV02\_4 (SEQ ID NO: 896)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV03\_4ASS (SEQ ID NO: 897)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV04\_4 (SEQ ID NO: 898)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV05\_4 (SEQ ID NO: 899)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV06\_4 (SEQ ID NO: 900)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTAHKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV07\_4 (SEQ ID NO: 901)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSL EEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAK\*

>ZV08\_4 (SEQ ID NO: 1107)

MKTFFKTL SAAALALILAACGGQKDSAPAASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ



VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV09\_4 (SEQ ID NO: 902)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV10\_4 (SEQ ID NO: 903)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV11\_4 (SEQ ID NO: 904)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV12\_4ASS (SEQ ID NO: 905)

MKTFFKTLASAAALALILAACGGQKDRAPAASASAAASENGAAKKEILFGTTVGDLGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV13\_4 (SEQ ID NO: 906)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV15\_4 (SEQ ID NO: 907)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAGNEGAAG\*

>ZV16\_4 (SEQ ID NO: 908)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV17\_4 (SEQ ID NO: 909)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV18\_4 (SEQ ID NO: 910)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVKLVEFTDYVRPNLALAEGLDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV19\_4 (SEQ ID NO: 911)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV20\_4 (SEQ ID NO: 912)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV21\_4 (SEQ ID NO: 913)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV22\_4 (SEQ ID NO: 914)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDLVKE  
QIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV24\_4ASS (SEQ ID NO: 915)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEQNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV25\_4 (SEQ ID NO: 916)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
QIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV26\_4 (SEQ ID NO: 917)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV27\_4 (SEQ ID NO: 918)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV28\_4 (SEQ ID NO: 919)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAEKKEIVFGTTVGDFGDMVKE  
HIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLSLEEVDKGSTVSAPNDPSNFARVLVMDDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNENGAAK\*

>ZV29\_4 (SEQ ID NO: 920)  
MKTFFKTLSSAAALALILAACGGQKDSAPAASASAAAADNGAAKKEIVFGTTVGDFGDMVKE  
QIQPELEKKGYTVLVEFTDYVRPNLALAEGLDINVQHKPYLDDFKKEHNLDITEVFQ

VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV32\_4 (SEQ ID NO: 921)

MKTFFKTLASAAALALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK  
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEAF  
QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS  
KADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV33\_4 (SEQ ID NO: 922)

MKTFFKTLASAAALALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK  
EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEAF  
QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS  
KADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW  
SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

>ZV96\_4 (SEQ ID NO: 923)

MKTFFKTLASAAALALILAACGGQKDSAPAASASAAADNGAEKKEIVFGTTVGDFGDMVKE  
QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ  
VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK  
ADIAENLNKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS  
AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKSPAAWNEGAAG\*

Figure 8 shows the results of aligning the sequences of each of these strains. Dark shading indicates regions of homology, and gray shading indicates the conservation of amino acids with similar characteristics. As is readily discernible, there is significant conservation among the various strains of ORF 4 (SEQ ID NO: 216), further confirming its utility as an antigen for both vaccines and diagnostics.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.